

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, C07K 14/47, A61K 38/17, G01N 33/68, C12Q 1/68, G01N 33/566, C07K 16/18, C12N 15/62</b>	<b>A2</b>	(11) International Publication Number: <b>WO 98/12327</b> (43) International Publication Date: 26 March 1998 (26.03.98)
--	-----------	--

(21) International Application Number: PCT/US97/16842  
(22) International Filing Date: 19 September 1997 (19.09.97)  
(30) Priority Data:  
60/025,296 20 September 1996 (20.09.96) US  
60/042,611 3 April 1997 (03.04.97) US  
60/042,985 4 April 1997 (04.04.97) US

(60) Parent Applications or Grants

(63) Related by Continuation

US	60/025,296 (CIP)
Filed on	20 September 1996 (20.09.96)
US	60/042,611 (CIP)
Filed on	3 April 1997 (03.04.97)
US	60/042,985 (CIP)
Filed on	4 April 1997 (04.04.97)

(71) Applicant (for all designated States except US): BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).-

(72) Inventors; and

(75) Inventors/Applicants (for US only): BOWCOCK, Anne, M. [GB/US]; 5823 Lupton Drive, Dallas, TX 75225 (US). BAER, Richard [US/US]; 2738 Gold Rush Lane, Carrollton, TX 75007 (US).

(74) Agent: HIBLER, David, W.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**

*Without international search report and to be republished upon receipt of that report.*

(54) Title: COMPOSITIONS AND METHODS COMPRISING BARD1 AND OTHER BRCA1 BINDING PROTEINS

(57) Abstract

Disclosed are several novel genes, identified in screening assays based upon binding to the breast cancer protein, BRCA1. The currently preferred gene and protein, termed BARD1, is a RING protein that interacts with BRCA1. The genes, proteins and other biological materials provided are envisioned for use in various cancer-related diagnostic and therapeutic methods, particularly those connected with breast, ovarian and uterine cancer.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## **BACKGROUND OF THE INVENTION**

### **COMPOSITIONS AND METHODS COMPRISING BARD1 AND OTHER BRCA1 BINDING PROTEINS**

5 The present application claims the priority of co-pending U.S. Provisional Patent Applications Serial No. 60/025,296, filed September 20, 1996, Serial No. 60/042,611, filed April 3, 1997, and Serial No. 60/042,985, filed April 4, 1997, the entire disclosures of which are incorporated herein by reference without disclaimer.

#### **1. Field of the Invention**

10

The present invention relates generally to the field of cancer, and particularly concerns the diagnosis and treatment of breast cancer. The invention provides novel genes, proteins and related compositions that interact with the BRCA1 gene product, which is known to be connected with a significant number of breast cancers. The currently preferred gene and protein  
15 of the invention is a RING protein termed BARD1. Also disclosed are various diagnostic and therapeutic methods and screening assays using the compositions of the invention.

#### **2. Description of Related Art**

20

Breast cancer is the most common fatal malignancy affecting women in the western world. The etiology of breast cancer is complex, and likely involves genetic, hormonal, environmental and other factors. Detailed analyses of breast cancer patients has revealed several alterations in gene expression associated with the disease. In addition to gene amplification, breast tumor development is thought to be the consequence of mutations in one or more  
25 recessive genes.

30

A particular breast cancer-related gene is the *BRCA1* gene. Germline mutations of the *BRCA1* gene are found in approximately half of families that display a heritable susceptibility to breast cancer (Hall *et al.*, 1990; Miki *et al.*, 1994; Futreal *et al.*, 1994; Castilla *et al.*, 1994; Simard *et al.*, 1994; Friedman *et al.*, 1994). In women of these kindreds, the mutant *BRCA1* allele confers lifetime risks of 80-90% for breast cancer and 40-50% for ovarian cancer (Easton

*et al.*, 1993; Ford *et al.*, 1994). The wild-type allele of *BRCA1* is typically lost or inactivated in the tumors that arise in these families, implying that *BRCA1* normally functions as a tumor-suppressor gene. A variety of different germline *BRCA1* mutations that segregate with breast cancer susceptibility have been described; these include missense mutations which produce  
5 single amino acid substitutions and, more commonly, frame-shifts or nonsense mutations which truncate the *BRCA1* reading frame (Miki *et al.*, 1994; Futreal *et al.*, 1994; Castilla *et al.*, 1994; Simard *et al.*, 1994; Friedman *et al.*, 1994).

The human *BRCA1* gene encodes a large polypeptide of 1863 amino acids, the precise  
10 biochemical function of which is not yet known (Miki *et al.*, 1994). A prominent feature of the protein is a RING domain that resides near its amino-terminus (residues 20-68). The RING motif, a cysteine-rich sequence found in a diverse group of regulatory proteins, adopts an interleaved structure in which two ions of zinc are coordinated by eight conserved amino acids (seven cysteines and one histidine) (Saurin *et al.*, 1996). Thus, *BRCA1* can be said to have two  
15 "zinc finger domains".

It has been proposed that the zinc fingers or RING domain serves as an interface for DNA recognition or protein-protein interactions (Saurin *et al.*, 1996), and that the *BRCA1* protein may be a transcription factor (Miki *et al.*, 1994; Vogelstein and Kinzler, 1994).  
20 However, no direct evidence that *BRCA1* is a transcription factor has yet been presented. In fact, a detailed characterization of *BRCA1* function at the molecular level has been somewhat hindered by the lack of purified protein in amounts sufficient to conduct productive assays *in vitro*.

25 Whatever its precise function, the analysis of germline mutations in families prone to breast and ovarian cancer suggests that the RING domain may be essential for the tumor suppressor activity of *BRCA1*; thus, in some kindreds the tumorigenic lesion is a single missense mutation (C61G or C64G) that specifically replaces one of the cysteine residues required for zinc coordination by the RING domain (Castilla *et al.*, 1994; Friedman *et al.*, 1994).

30 Recent studies have shown that the mouse and human homologs of *BRCA1* share approximately 60% amino acid identity (Bennett *et al.*, 1995; Lane *et al.*, 1995; Sharan *et al.*,



1995). This degree of phylogenetic conservation is low, especially when compared with other known tumor suppressor proteins; for example, the mouse and human counterparts of RB1, p53, APC, WT1, and NF1 display amino acid identities in the range of 78-98%. Nevertheless, two regions of BRCA1 are especially well conserved. The first corresponds to the amino-terminal 100 residues; this sequence encompasses the RING domain and the tumorigenic missense mutations at C61 and C64 (Castilla *et al.*, 1994; Friedman *et al.*, 1994).

The second region of high conservation resides near the carboxy-terminus of BRCA1, and it also serves as a target for missense mutations associated with familial breast cancer (Sharan *et al.*, 1995). This region includes two tandem copies of the BRCA1 carboxy-terminal domain ("BRCT domain"), a newly-recognized amino acid motif also found in 53BP1, a mammalian polypeptide that binds the p53 tumor suppressor, and RAD9, a yeast protein that mediates cell cycle arrest in response to DNA damage (Koonin *et al.*, 1996).

Given that the BRCA1 gene and protein product are now accepted to be closely linked to familial breast cancer development, but that the function of BRCA1 remains unknown, any further delineation of the properties and interactions of the BRCA1 protein would be an important development. The identification of proteins that bind to BRCA1 would be particularly beneficial as they themselves would likely be implicated in the breast cancer process. The cloning of genes encoding such BRCA1-binding proteins would therefore be a significant contribution towards the development of further cancer diagnostics and therapeutics.

### SUMMARY OF THE INVENTION

The present invention provides several novel genes, proteins and related biological compositions developed from their ability to bind to the BRCA1 protein. Methods of using the various compositions, for example, in the diagnosis, prognosis and treatment of breast, ovarian and uterine cancer are also provided.

The present invention first provides DNA segments, vectors and the like comprising at least a first isolated gene, DNA segment or coding sequence region that encodes a BARD1,

B123, BE2, BE14, BE31 or BE445 protein, polypeptide, domain, peptide or any fusion protein thereof, and particularly, that encode a human BARD1, B123, BE2, BE14, BE31 or BE445 protein, domain, fragment or derivative.

5           As used herein in the context of the instant compositions, the term BARD1, B123, BE2, BE14, BE31 and BE445 will be understood to include wild-type, polymorphic and mutant BARD1, B123, BE2, BE14, BE31 and BE445 sequences. Wild-type sequences are defined as the first identified sequence, polymorphic sequences are defined as naturally occurring variants of the wild-type sequence that have no effect on the expression or function of the BARD1,  
10   B123, BE2, BE14, BE31 or BE445 proteins or domains thereof, and mutant sequences are defined as changes in the wild-type sequence, either naturally occurring or introduced by the hand of man, that have an effect on either the expression and/or the function of the BARD1, B123, BE2, BE14, BE31 or BE445 proteins or domains thereof.

15           Thus, the invention also includes the provision of DNA segments, vectors, genes and coding sequence regions that encode BARD1, B123, BE2, BE14, BE31 or BE445 proteins, polypeptides, domains, peptides or any fusion protein thereof, where the BARD1, B123, BE2, BE14, BE31 or BE445 protein element comprises at least one mutation in comparison to the wild-type sequence. The mutation may be deliberately introduced by the hand of man, for  
20   example, in order to test the function of the changed amino acid, e.g., in BRCA1 binding, DNA binding and/or other functions. Additionally, the mutation may be a naturally occurring polymorphic change, either isolated from normal cells or introduced by the hand of man.

25           The BARD1, B123, BE2, BE14, BE31 or BE445 mutation may also be in a purified protein obtained directly from an aberrant cell, such as a breast, ovarian or uterine cancer cell, or may be a recombinant protein that has been changed to introduce a mutation that mirrors one identified in a patient. The mutation may result in a truncated BARD1, B123, BE2, BE14, BE31 or BE445 gene or protein, or may result in increased, decreased or undetectable levels of BARD1, B123, BE2, BE14, BE31 or BE445 gene or protein being produced. Where diagnostic  
30   or prognostic mutated BARD1, B123, BE2, BE14, BE31 or BE445 genes, proteins and antibodies are concerned the mutant gene, DNA segment, antibody or even peptide will preferably have specificity for the mutant sequence in preference to the wild-type sequence,

allowing effective differentiation between the two, as may be used in diagnostic or prognostic tests for breast, ovarian or uterine cancer cells or patients, as described in more detail herein below.

5           The DNA segments and vectors may comprise an isolated gene or coding sequence that encodes a BARD1 protein characterized as having the following properties:

being about 777, 770 or about 752 amino acids in length, preferably being 777 amino acids in length;

10           comprising an amino-terminal RING motif or domain, preferably characterized as comprising a cysteine-rich sequence with an interleaved structure in which two ions of zinc are coordinated by seven cysteines and one histidine, and which RING motif or domain mediates the association of BARD1 with BRCA1;

containing ankyrin repeats, which ankyrin repeats are not required for binding to BRCA1;

15           comprising carboxy-terminal BRCT domains that are homologous to carboxy-terminal sequences of BRCA1;

being encoded by sequences on chromosome 2q;

20           binding to BRCA1, as may be assessed by one or more cellular assay systems, such as a yeast or mammalian two-hybrid system that identifies functional proteins associations *in vivo*; or by co-immunoprecipitation of the BRCA1 and BARD1 proteins from mammalian cell lysates, or by using one or more *in vitro* assays of protein binding;

25           and more preferably, characterized as binding to the amino-terminal region of BRCA1, most preferably to the BRCA1 amino-terminal 101 residues that encompasses the RING motif (residues 20-68), but as not binding to the BRCA1 fragment between residues 1 and 71;

and even more preferably, wherein residues 26-202 of BARD1, and most preferably, where residues 26-142 of BARD1, which include the RING motif (residues 46-90), but do not include the ankyrin repeats (residues 427-525), interact with BRCA1.

5

It will be understood that while the normal, native, wild-type BARD1 protein is defined in terms of these properties and domains, the overall features will generally be the same for BARD1 polymorphic and mutant proteins and domains as well. The polymorphic and mutant BARD1 genes and proteins can be understood with reference to the wild-type sequences and the exemplary mutants included herein.

10

The genes and DNA segments of the present invention preferably encode wild-type or polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BARD1 sequence includes a contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 or SEQ ID NO:39, or a biologically functional equivalent thereof. The present invention also provides genes and DNA segments that encode mutant BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BARD1 sequence includes a contiguous amino acid sequence from SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37, or a biologically functional equivalent thereof. As used herein, the term "contiguous amino acid sequence" will be understood to include a contiguous amino acid sequence of at least about 4, about 6, about 9, about 10, about 12, about 15 or about 20 amino acids or so.

15

20

25

Thus in certain aspects of the present invention, the genes and DNA segments encode wild-type BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the wild-type BARD1 sequence includes a contiguous amino acid sequence from SEQ ID NO:2 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:1 or a biologically functional equivalent thereof.

30

In other aspects of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof

where the polymorphic BARD1 sequence is described as BARD1 P143, and includes a contiguous amino acid sequence from SEQ ID NO:21 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:20 or a biologically functional equivalent thereof.

In further embodiments of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the polymorphic BARD1 sequence is described as BARD1 P531, and includes a contiguous amino acid sequence from SEQ ID NO:23 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:22 or a biologically functional equivalent thereof.

In yet other aspects of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the polymorphic BARD1 sequence is described as BARD1 P1121, and includes a contiguous amino acid sequence from SEQ ID NO:25 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:24 or a biologically functional equivalent thereof.

In still other embodiments of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the polymorphic BARD1 sequence is described as BARD1 PΔ1140-1160, and includes a contiguous amino acid sequence from SEQ ID NO:27 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2385 of SEQ ID NO:26 or a biologically functional equivalent thereof.

In alternate aspects of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof

where the polymorphic BARD1 sequence is described as BARD1 P1592, and includes a contiguous amino acid sequence from SEQ ID NO:29 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:28 or a biologically functional equivalent thereof.

In particular embodiments of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the polymorphic BARD1 sequence is described as BARD1 P1765, and includes a contiguous amino acid sequence from SEQ ID NO:31 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:30 or a biologically functional equivalent thereof.

In particular embodiments of the present invention, the genes and DNA segments encode polymorphic BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the polymorphic BARD1 sequence is described as BARD1 P2354, and includes a contiguous amino acid sequence from SEQ ID NO:39 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:38 or a biologically functional equivalent thereof.

In certain embodiments of the present invention, the genes and DNA segments encode mutant BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the mutant BARD1 sequence is described as BARD1 MQ564H, and includes a contiguous amino acid sequence from SEQ ID NO:33 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:32 or a biologically functional equivalent thereof.

In other aspects of the present invention, the genes and DNA segments encode mutant BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the mutant

BARD1 sequence is described as BARD1 MS761N, and includes a contiguous amino acid sequence from SEQ ID NO:35 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:34 or a biologically functional equivalent thereof.

5

In further embodiments of the present invention, the genes and DNA segments encode mutant BARD1 proteins, polypeptides, domains, peptides or fusion constructs thereof where the mutant BARD1 sequence is described as BARD1 MR658C, and includes a contiguous amino acid sequence from SEQ ID NO:37 or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 75 and position 2405 of SEQ ID NO:36 or a biologically functional equivalent thereof.

10

The DNA segments and coding regions may encode wild-type, polymorphic or mutant BARD1 peptides, *e.g.*, of from about 15 to about 30 or about 50 amino acids in length or so. The BARD1 peptides may be lacking in any defined BARD1 activity, and may, for example, be used in generating antibodies or in other embodiments. The BARD1 peptides or domains may also be deliberately engineered to include a mutation, *e.g.*, in order to prepare antibodies that are specific for a mutated BARD1, particularly where the mutation represents one identified in a patient with breast, ovarian or endometrial cancer.

15

20

The present invention also provides DNA segments and coding regions that may encode a BARD1 peptide of from about 6 to about 30 amino acids in length, the peptide having an amino acid sequence that corresponds to a wild-type BARD1 sequence of a BARD1 protein sequence region that is susceptible to mutations that are indicative of a malignant phenotype. Where diagnostic or prognostic BARD1 genes, proteins and antibodies are concerned the gene, DNA segment, antibody or even peptide will preferably allow effective differentiation between the mutant BARD1 sequence and the wild-type BARD1 sequence as may be used in diagnostic or prognostic tests for breast, ovarian or uterine cancer cells or patients, as described in more detail herein below.

25

30

The genes, DNA segments, vectors and coding sequence regions may also encode wild-type, polymorphic or mutant BARD1 polypeptides and peptides with certain, but necessary all, BARD1 functional properties. As such genes and coding sequences encoding isolated wild-type, polymorphic or mutant BARD1 domains are provided.

5

The wild-type, polymorphic or mutant BARD1 domains contemplated include isolated and/or purified wild-type, polymorphic or mutant BARD1 ankyrin repeat domains, including those comprising three ankyrin repeats and comprising or having the sequence of residues 427-525 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39; isolated and/or purified BARD1 BRCT-like domains, as exemplified by those comprising the BRCT domain N-terminal core motif of residues 616-653 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, the BRCT domain C-terminal core motif of residues 743-777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, the BRCT domain of residues 616-777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39; and isolated and/or purified BARD1 RING motif domains exemplified by those comprising or having the sequence of residues 46-90 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39.

Preferred examples of domains are the BRCA1 binding domains. For example, those comprising or having the sequence of residues 26-202 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, or more preferably, those comprising or having the sequence of residues 26-142 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, or any active portion of such sequences that functions to bind BRCA1.

30



"BRCA1 binding", as used herein, may be assessed by any one or more suitable *in vitro*, *in vivo* or *in cellulo* assays. For example, co-immunoprecipitation of the BRCA1 and BARD1 proteins from mammalian cell lysates, and *in vitro* assays of protein binding, *e.g.*, wherein one or both of the BARD1 or BRCA1 components are attached to a detectable label, and/or are immobilized may be employed. Cellular assay systems, such as a yeast or mammalian two-hybrid protein association system may also be employed, as disclosed herein.

The BARD1 domains may also be mutant domains, which include naturally occurring polymorphisms, mutations found in BARD1 proteins in patients and, also, mutations deliberately engineered into a domain to test their function in assays. The mutant domains are also useful in antibody generation and in various *in vitro* and cellular assays. Engineering increased BRCA1 binding is also contemplated.

The full length wild-type, polymorphic and mutant BARD1 proteins of the present invention are unusual in that they combine sequence features and motifs not previously observed in combination, *e.g.*, RING and BRCT elements. The wild-type, polymorphic and mutant BARD1 proteins of the invention may be further characterized as including domains defined as:

comprising an amino-terminal RING motif or domain that has the sequence of residues 46-90 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39;

comprising a binding domain, or "BRCA1 binding domain" that has the sequence of residues 26-202 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, or more preferably, that has the sequence of residues 26-142 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, which binding domain binds to BRCA1;

containing ankyrin repeats that have the sequence of residues 427-525 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, which ankyrin repeats do not bind to BRCA1; and

comprising carboxy-terminal BRCT domains that have a sequence between residues 605 and 777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, as exemplified by comprising the BRCT domain N-terminal core motif of residues 616-653 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39 and as comprising the BRCT domain C-terminal core motif of residues 743-777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39.

As the full length DNA segments of the invention preferably encode wild-type, polymorphic or mutant BARD1 proteins of about 777, 770 or 752 amino acids in length, each of the sequence designations provided herein refer to the 777, 770 or 752 amino acid sequence of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39. However, with proteins of shorter length, the operative domains and regions will be easily identified by virtue of the sequence and respective locations.

DNA segments, isolated genes or coding regions may also be manipulated to encode BARD1, B123, BE2, BE14, BE31 or BE445 fusion proteins or constructs in which at least one BARD1, B123, BE2, BE14, BE31 or BE445 protein sequence is operatively attached or linked to at least one distinct, selected amino acid sequence. The combination of BARD1, B123, BE2, BE14, BE31 or BE445 sequences with selected antigenic amino acid sequences; selected non-antigenic carrier amino acid sequences, for use in immunization; selected adjuvant sequences;

amino acid sequences with specific binding affinity for a selected molecule; and amino acid sequences that form an active DNA binding or transactivation domain are particularly contemplated. Certain fusion proteins may be linked together via a protease-sensitive peptide linker, allowing subsequent easy separation.

5

Also particularly contemplated are the combination of BARD1, B123, BE2, BE14, BE31 or BE445 sequences with a selected tumor suppressor protein or peptide. Tumor suppressor proteins contemplated for use include, but are not limited to, the retinoblastoma, p53, Wilms tumor (WT-1), DCC, neurofibromatosis type 1 (NF-1), von Hippel-Lindau (VHL) disease tumor suppressor, Maspin, Brush-1, BRCA-1, BRCA-2 and the multiple tumor suppressor (MTS) or p16 proteins or peptides. Further particularly contemplated are the combination of BARD1, B123, BE2, BE14, BE31 or BE445 sequences with a selected wild-type version of a selected oncogenic protein or peptide. Wild-type oncogenic proteins contemplated for use include, but are not limited to, tyrosine kinases, both membrane-associated and cytoplasmic forms, such as members of the Src family, serine/threonine kinases, such as Mos, growth factor and receptors, such as platelet derived growth factor (PDGF), small GTPases (G proteins) including the ras family and Gs-alpha, cyclin-dependent protein kinases (cdk), members of the myc family members including c-myc, N-myc, and L-myc and bcl-2 and family members.

10

15

20

DNA segments and isolated genes may also be manipulated to encode BARD1, B123, BE2, BE14, BE31 or BE445 fusion proteins or constructs in which at least one BARD1, B123, BE2, BE14, BE31 or BE445 protein sequence is operatively attached or linked to at least one distinct, selected BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide sequence.

25

The DNA segments intended for use in expression will be operatively positioned under the control of, *i.e.*, downstream from, a promoter that directs expression of BARD1, B123, BE2, BE14, BE31 or BE445 in a desired host cell, such as *E. coli*, or in certain preferred embodiments in a mammalian or human cell. The promoter may be a recombinant promoter or a promoter naturally associated with a BARD1, B123, BE2, BE14, BE31 or BE445 gene. Recombinant vectors thus form another aspect of the present invention.

30

The use of isolated BARD1, B123, BE2, BE14, BE31 or BE445 genes positioned, in reverse orientation, under the control of a promoter that directs the expression of an antisense product in a cell is also contemplated.

5 In certain aspects of the present invention, the nucleic acid segments disclosed herein further comprise a second sequence region of at least about 20 contiguous nucleotides that have the same sequence as, or are complementary to, SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ  
10 ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, said sequence region and said second sequence region from spatially distant regions within SEQ ID NO:1, SEQ ID NO:9, SEQ  
15 ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ  
20 ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130.

In the same yeast two-hybrid system used to identify BARD1, fourteen other novel genes that encode polypeptides that bind to BRCA1 were identified. These are the TCL52 DNA and protein sequence (SEQ ID NO:9 and SEQ ID NO:48, respectively); TCL163 DNA and protein  
25 sequence (SEQ ID NO:10 and SEQ ID NO:49, respectively); B223 DNA and protein sequence (SEQ ID NO:11 and SEQ ID NO:50, respectively); B115 DNA and protein sequence (SEQ ID NO:12 and SEQ ID NO:51, respectively); BAP28 DNA and protein sequence (SEQ ID NO:13 and SEQ ID NO:52, respectively); B48 DNA and protein sequence (SEQ ID NO:14 and SEQ ID NO:53, respectively); B258 DNA and protein sequence (SEQ ID NO:15 and SEQ ID NO:54,  
30 respectively); BAP152 DNA and protein sequence (SEQ ID NO:16 and SEQ ID NO:55, respectively); B123 DNA and protein sequence (SEQ ID NO:17 and SEQ ID NO:19, respectively); B268 DNA and protein sequence (SEQ ID NO:18 and SEQ ID NO:56,

respectively); BE2 DNA and protein sequence (SEQ ID NO:40 and SEQ ID NO:41, respectively); BE14 DNA and protein sequence (SEQ ID NO:42 and SEQ ID NO:43, respectively); BE31 DNA and protein sequence (SEQ ID NO:44 and SEQ ID NO:45, respectively); and BE445 DNA and protein sequence (SEQ ID NO:46 and SEQ ID NO:47, respectively).

Thus, the present invention further advantageously provides methods for identifying a human candidate tumor suppressor gene or oncogene based upon the "two hybrid screening system". One such method may be characterized as comprising the steps of:

- a) obtaining a first DNA segment comprising a candidate human gene; the first DNA segment expressing a first fusion protein comprising a transcriptional transactivating domain operatively attached to the candidate protein encoded by the candidate gene;
- b) obtaining a second DNA segment that expresses a second fusion protein comprising a human BRCA1 or BARD1 RING domain operatively attached to a DNA binding domain that binds to a defined nucleic acid sequence;
- c) providing the first and second DNA segments to a eukaryotic host cell that comprises a marker gene operatively positioned downstream of the defined nucleic acid sequence; and
- d) identifying a eukaryotic host cell that expresses the marker gene, thereby identifying the candidate gene as a human gene that encodes a tumor suppressor gene or oncogene.

The methods generally further comprise isolating the identified candidate human tumor suppressor gene or oncogene from the first DNA segment within the eukaryotic host cell.

The transcriptional transactivating domains used in the present invention may be the GAL4, HAP1, LEU3, PHO4, PHO2, PPR1, ARGRII, ADR1, QA1F, MAL63, LAC9, GCN4 or

VP16 transcriptional transactivating domain. The fusion protein may comprise a GAL4 DNA binding domain, wherein the defined nucleic acid sequence comprises a GAL4 binding domain recognition sequence, or a lexA DNA binding domain, wherein the defined nucleic acid sequence comprises a lexA binding site sequence. In the methods, the eukaryotic host cell may be a yeast host cell (yeast two hybrid system) or a mammalian host cell.

In the two hybrid system methods of the present invention, marker genes preferred for use are chloramphenicol acetyltransferase,  $\beta$ -galactosidase, green fluorescent protein,  $\beta$ -glucuronidase or the luciferase gene, preferably the  $\beta$ -galactosidase gene. In other aspects, the marker genes can be genes that encode vital biological components, used in combination with strains of *Saccharomyces cerevisiae* that lack one or more of these genes, such that expression of one or more of the marker genes is required to produce viable colonies. Marker genes contemplated for use in these aspects of the invention are exemplified by, but not limited to, the *URA3*, *TRP1*, *HIS3*, *LYS2*, *ADE1* and *LEU2* genes of *Saccharomyces cerevisiae*.

A further explanation of the two hybrid system cloning method for identifying a human gene that encodes a candidate tumor suppressor protein or oncogene is that it generally operatively comprises the steps of:

- a) obtaining a plurality of first DNA segments comprising a plurality of candidate human genes;
- b) obtaining multiple copies of the second DNA segment;
- c) providing the plurality of first DNA segments and multiple copies of the second DNA segments to a population of eukaryotic host cells in an amount sufficient to provide about one first DNA segment and at least about one second DNA segment to each host cell in the population;
- d) culturing the population of cells under conditions and for a period of time effective to allow marker gene expression; and

- e) detecting a host cell from the population that expresses the marker gene, thereby identifying the presence in the cell of a first DNA segment that comprises a candidate tumor suppressor protein or oncogene.

5 In a preferred method of the present invention, the plurality of candidate human genes are the plurality of genes in a B-cell, breast, ovarian or uterine DNA library. The method also generally further comprises isolating the detected cell of step (e) free from the population of cells, and isolating the candidate human gene from the first DNA segment within the cell.

10 The genes and DNA segments of the present invention may encode B123 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B123 sequence includes a contiguous amino acid sequence from SEQ ID NO:19, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 46 and position 864 of SEQ ID NO:17, or a biologically  
15 functional equivalent thereof.

The genes and DNA segments of the present invention may encode BE2 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BE2 sequence includes a contiguous amino acid sequence from SEQ ID NO:41, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid  
20 sequence from between position 37 and position 819 of SEQ ID NO:40, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode BE14 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BE14 sequence includes a contiguous amino acid sequence from SEQ ID NO:43, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 666 of SEQ ID NO:42, or a biologically  
25 functional equivalent thereof.

30

The genes and DNA segments of the present invention may encode BE31 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BE31 sequence includes

a contiguous amino acid sequence from SEQ ID NO:45, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 693 of SEQ ID NO:44, or a biologically functional equivalent thereof.

5

The genes and DNA segments of the present invention may encode BE445 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BE445 sequence includes a contiguous amino acid sequence from SEQ ID NO:47, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 816 of SEQ ID NO:46, or a biologically functional equivalent thereof.

10

The genes and DNA segments of the present invention may encode TCL52 proteins, polypeptides, domains, peptides or fusion constructs thereof where the TCL52 sequence includes a contiguous amino acid sequence from SEQ ID NO:48, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 936 of SEQ ID NO:9, or a biologically functional equivalent thereof.

15

The genes and DNA segments of the present invention may encode TCL163 proteins, polypeptides, domains, peptides or fusion constructs thereof where the TCL163 sequence includes a contiguous amino acid sequence from SEQ ID NO:49, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 7 and position 1770 of SEQ ID NO:10, or a biologically functional equivalent thereof.

20

25

The genes and DNA segments of the present invention may encode B223 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B223 sequence includes a contiguous amino acid sequence from SEQ ID NO:50, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 1110 of SEQ ID NO:11, or a biologically functional equivalent thereof.

30



The genes and DNA segments of the present invention may encode B115 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B115 sequence includes a contiguous amino acid sequence from SEQ ID NO:51, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 1248 of SEQ ID NO:12, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode BAP28 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BAP28 sequence includes a contiguous amino acid sequence from SEQ ID NO:52, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 1545 of SEQ ID NO:13, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode B48 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B48 sequence includes a contiguous amino acid sequence from SEQ ID NO:53, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 3 and position 449 of SEQ ID NO:14, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode B258 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B258 sequence includes a contiguous amino acid sequence from SEQ ID NO:54, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 1 and position 1605 of SEQ ID NO:15, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode BAP152 proteins, polypeptides, domains, peptides or fusion constructs thereof where the BAP152 sequence includes a contiguous amino acid sequence from SEQ ID NO:55, or a biologically functional

equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 959 and position 2143 of SEQ ID NO:16, or a biologically functional equivalent thereof. Alternatively, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 2147 and position 2605 of SEQ ID NO:16, or a biologically functional equivalent thereof.

The genes and DNA segments of the present invention may encode B268 proteins, polypeptides, domains, peptides or fusion constructs thereof where the B268 sequence includes a contiguous amino acid sequence from SEQ ID NO:56, or a biologically functional equivalent thereof. Preferably, the isolated genes and coding regions will include a contiguous nucleic acid sequence from between position 46 and position 864 of SEQ ID NO:18, or a biologically functional equivalent thereof.

The nucleic acid segments provided by the invention are thus further characterized as including:

(a) a nucleic acid segment comprising a sequence region that consists of at least about 8, about 10, about 11, about 12, about 13, about 14, about 15, about 17 or about 20 contiguous nucleotides that have the same sequence as, or are complementary to, about 8, about 10, about 11, about 12, about 13, about 14, about 15, about 17 or about 20 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130; or

(b) a nucleic acid segment of from about 10-14, 17 or about 20 to about 20,000 nucleotides in length that specifically hybridizes to the nucleic acid segment of

SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, or the complements thereof, under standard stringency, or preferably, under high stringency hybridization conditions.

Standard and high stringency hybridization conditions are well known to those of skill in the art. An exemplary, but not limiting, standard hybridization is incubated at 42°C in 50% formamide solution containing dextran sulfate for 48 hours and subjected to a final wash in 0.5X SSC, 0.1% SDS at 65°C. In addition to hybridization to Southern or northern blots, hybridization of primers for use in PCR™, as exemplified in Example XI below, is another preferred method for identification of sequences contemplated for use in the present invention.

Where the "complement" of any of the above nucleic acid segments are provided, such a complement may be functionally considered as an antisense nucleic acid, which includes nucleic acid segments positioned, in reverse orientation, under the control of a promoter that directs the expression of an antisense product. Antisense products may be used to inhibit the transcription or translation of any of the foregoing BRCA1-binding genes, in *in vitro* systems in order to more precisely define the cellular consequence of inhibition, or even *in vivo* in situations where inhibition of one or more of the foregoing BRCA1-binding genes would be believed to result in a beneficial effect, such as an anti-cancer effect.

Mutants of each of the foregoing sequences and their encoded proteins, polypeptides, and peptides are also contemplated. The mutants may be used in the detection of physiologically relevant mutations or in further testing and functional analyses.

Segments of each of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID

NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, or the complements thereof, or the mutants thereof, may variously be about 10, 14, 17, 20, 25, 30, 50, 100, 200, 500, or 1000 or so nucleotides in length, up to and including the full length sequences, or even longer, as may be achieved by duplication of certain domains. Where the wild-type, polymorphic or mutant BARD1 sequences of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, or SEQ ID NO:38 are concerned, sequences of at least about 1500 or about 2000 nucleotides of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, or SEQ ID NO:38, or the complement thereof are provided, up to and including the full length sequence of 2531 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, or SEQ ID NO:38, or up to and including the full length sequence of 2510 contiguous nucleotides of SEQ ID NO:26, or the complement thereof.

Any segment may be combined into a DNA segment or vector of up to about 50,000, about 30,000, or about 20,000 basepairs in length. Segments of up to about 20,000, 15,000 or about 10,000 basepairs in length will generally be preferred, and segments of up to about 5,000 and 3,000 basepairs in length are also provided.

The nucleic acids of the present invention may also be DNA segments or RNA segments. Nucleic acid detection kits are also provided.

The present invention further provides recombinant host cells comprising at least one DNA segment or vector that comprises an isolated gene that encodes a BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, domain, peptide or any fusion protein or mutant thereof. Prokaryotic recombinant host cells, such as *E. coli*, are provided, as are eukaryotic host

cells, including breast, ovarian or uterine cancer cells provided with the BARD1, B123, BE2, BE14, BE31 or BE445 constructs of the invention.

The recombinant host cells may further comprise an operative BRCA1 protein or active  
5 fragment or domain thereof, such as a DNA binding domain and/or a BARD1, B123, BE2, BE14, BE31 or BE445 binding domain. Such recombinant host cells may be provided with the BRCA1 *in vitro*, for example, to test BARD1, B123, BE2, BE14, BE31 or BE445 and BRCA1 interactions, or may naturally express BRCA1, including cells provided with BARD1, B123, BE2, BE14, BE31 or BE445 *in vivo* and *in vitro*, either for treatment or for study.

10 The recombinant host cells of the present invention preferably have one or more DNA segments introduced into the cell by means of a recombinant vector, and preferably express the DNA segment to produce the encoded BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide.

15 Methods of using BARD1, B123, BE2, BE14, BE31 or BE445 DNA segments are provided that comprise expressing a BARD1, B123, BE2, BE14, BE31 or BE445 DNA segment in a recombinant host cell and collecting the BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide, domain or mutant expressed by said cell. These methods may be characterized  
20 by the steps of:

- (a) preparing a recombinant vector in which a BARD1, B123, BE2, BE14, BE31 or BE445-encoding DNA segment is positioned under the control of a promoter;  
25
- (b) introducing said recombinant vector into a recombinant host cell;
- (c) culturing the recombinant host cell under conditions effective to allow expression of an encoded BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide, domain or mutant; and  
30

- (d) collecting said expressed BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide, domain or mutant.

Thus the present invention provides BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segments for use in the preparation of a recombinant BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, mutant or fusion protein thereof. Thus, the use of BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segments in the preparation of a recombinant BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, mutant or fusion protein thereof is provided.

Methods for detecting BARD1, B123, BE2, BE14, BE31 or BE445 genes in cells or samples are also provided and generally comprise contacting sample nucleic acids from a sample suspected of containing BARD1, B123, BE2, BE14, BE31 or BE445 with a nucleic acid segment that encodes a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide under conditions effective to allow hybridization of substantially complementary nucleic acids, and detecting the hybridized complementary nucleic acids thus formed.

The present invention also provides BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segments for use in the preparation of a composition for use in detecting a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segment. Thus, the use of BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segments in the preparation of a composition for use in detecting a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segment are provided. The invention further provides BARD1 nucleic acid segments for use in the preparation of a wild-type BARD1 composition for use in detecting or purifying a BRCA1 protein. Therefore, the use of BARD1 nucleic acid segments in the preparation of a wild-type BARD1 composition for use in detecting or purifying a BRCA1 protein is provided.

The methods may be diagnostic of breast, ovarian or uterine cancer by detecting BARD1, B123, BE2, BE14, BE31 or BE445 mutants as opposed to wild-type sequences. The use of both BARD1, B123, BE2, BE14, BE31 or BE445 wild-type and mutant sequences as probes or primers in such methods will naturally be included. A wild-type sequence probe or primer will be expected to bind to the native, non-mutant sequences, but not to a mutant, and

*vice versa*. The use of a mutant-specific probe that corresponds to a mutant identified in a family member with breast cancer may be preferred in screening other family members. In any event, irrespective of the BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segment employed, these studies will still only allow hybridization of substantially complementary  
5 nucleic acids, thus facilitating the detection only of wild-type or only mutant hybridized nucleic acid complexes.

Thus the present invention provides BARD1, B123, BE2, BE14, BE31 or BE445 compositions for use in the preparation of a diagnostic formulation for use in identifying a  
10 patient having or at risk for developing cancer. Therefore, the use of BARD1, B123, BE2, BE14, BE31 or BE445 compositions in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer is provided.

In further embodiments, the present invention provides BARD1, B123, BE2, BE14,  
15 BE31 or BE445 proteins, polypeptides, domains, peptides, mutants and any fusion proteins thereof, including BARD1, B123, BE2, BE14, BE31 or BE445 compounds purified from natural sources, such as from mammalian and human cells, and BARD1, B123, BE2, BE14, BE31 or BE445 prepared by recombinant means. Recombinant BARD1, B123, BE2, BE14, BE31 or BE445 proteins and peptides may be defined as being prepared by expressing a BARD1, B123,  
20 BE2, BE14, BE31 or BE445 protein or peptide in a recombinant host cell and purifying the expressed BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide away from total recombinant host cell components.

The BARD1, B123, BE2, BE14, BE31 or BE445 protein compositions, whether natural  
25 or recombinant, will generally be obtained free from total cell components, and will comprise at least one type of isolated BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide, purified relative to the natural level in a given cell.

As stated, preferred wild-type, polymorphic or mutant BARD1 proteins may be  
30 characterized as being about 777, about 770 or about 752 amino acids in length, preferably being 777 amino acids in length; as comprising an amino-terminal RING motif or domain, preferably characterized as comprising a cysteine-rich sequence with an interleaved structure in which two

ions of zinc are coordinated by seven cysteines and one histidine, and which RING motif or domain mediates the association of wild-type, polymorphic or mutant BARD1 with BRCA1; as containing ankyrin repeats, which ankyrin repeats are not required for binding to BRCA1; as comprising carboxy-terminal BRCT domains that are homologous to carboxy-terminal sequences of BRCA1; as being encoded by sequences on chromosome 2q; and most importantly in functional terms, as binding to BRCA1.

The wild-type, polymorphic or mutant BARD1 proteins of the invention are preferably characterized as comprising an amino-terminal RING motif or domain that has the sequence of residues 46-90 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39; as comprising a BRCA1 binding domain that has the sequence of residues 26-202 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, or more preferably, that has the sequence of residues 26-142 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, which binding domain binds to BRCA1; as containing ankyrin repeats that have the sequence of residues 427-525 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, which ankyrin repeats do not bind to BRCA1; and as comprising carboxy-terminal BRCT domains that have a sequence between residues 605 and 777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, as exemplified by comprising the BRCT domain N-terminal core motif of residues 616-653 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39 and as comprising the BRCT domain C-terminal core motif of residues 743-777 of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39.



Wild-type, polymorphic and mutant BARD1 domains and peptides are also provided by the invention, including the isolated wild-type, polymorphic or mutant BARD1 ankyrin repeat domains, isolated wild-type, polymorphic or mutant BARD1 BRCT-like domains, isolated wild-type, polymorphic or mutant BARD1 RING motif domains and the isolated wild-type, polymorphic or mutant BARD1 BRCA1-binding domains, and the non-functional antigenic peptides, as detailed hereinabove.

BARD1, B123, BE2, BE14, BE31 or BE445 fusion proteins or constructs including BARD1, B123, BE2, BE14, BE31 or BE445 sequences operatively attached to distinct, selected amino acid sequences, such as selected antigenic amino acid sequences, amino acid sequences with selected binding affinity, and DNA binding or transactivation amino acid sequences, are also encompassed within the invention. Fusion proteins with selectably-cleavable bonds are also provided.

The present invention provides BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains and fusion proteins for use in detection or purification of a BRCA1 protein. Thus, the use of BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains and fusion proteins in detection or purification of a BRCA1 protein is provided.

20

The BARD1, B123, BE2, BE14, BE31 or BE445 proteinaceous compositions will include the same types of mutants as described above for the nucleic acids. The use of specific mutated BARD1, B123, BE2, BE14, BE31 or BE445 peptides to prepare mutant-specific antibodies is particularly contemplated. In terms of diagnostic mutated BARD1, B123, BE2, BE14, BE31 or BE445 peptides and antibodies, these compositions will generally be more useful in regard to point mutants, whereas nucleic acid probes may be more suitable for detecting deletion, duplication, translocation and insertional mutations in addition to point mutants.

25

In still further embodiments, the present invention provides compositions comprising BARD1, B123, BE2, BE14, BE31 or BE445 in combination with an operative BRCA1 protein or active fragment or domain thereof. Such compositions may comprise BARD1, B123, BE2,

30

BE14, BE31 or BE445 in functional association with a BRCA1 protein or fragment, or may even comprise one or more BARD1, B123, BE2, BE14, BE31 or BE445-BRCA1 fusion proteins.

5           The BARD1, B123, BE2, BE14, BE31 or BE445 proteins, polypeptides, domains, peptides and fusion proteins, as well as the BARD1, B123, BE2, BE14, BE31 or BE445 DNA segments, vectors, isolated genes and coding sequences may also be formulated with a pharmaceutically acceptable diluent or vehicle to form a BARD1, B123, BE2, BE14, BE31 or BE445 pharmaceutical composition in accordance with this invention.

10           Further compositions of the present invention are antibodies, including monoclonal antibodies and antibody conjugates, that have immunospecificity for a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide. The antibodies may be operatively attached to a detectable label. The antibodies and antibody conjugates may be specific for mutant BARD1,  
15           B123, BE2, BE14, BE31 or BE445 proteins or peptides and allow differential binding from wild-type BARD1, B123, BE2, BE14, BE31 or BE445. Antibody detection kits are also provided.

20           Thus, the present invention provides BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains, mutants and fusion proteins thereof for use in the production of anti-BARD1, anti-B123, anti-BE2, anti-BE14, anti-BE31 and anti-BE445 antibodies. Therefore, the use of BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains, mutants and fusion proteins thereof in the production of anti-BARD1, anti-B123, anti-BE2, anti-BE14, anti-BE31 and anti-BE445 antibodies is provided.  
25           The anti-BARD1, anti-B123, anti-BE2, anti-BE14, anti-BE31 and anti-BE445 antibodies are also contemplated for use in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer. Thus, the use of anti-BARD1, anti-B123, anti-BE2, anti-BE14, anti-BE31 and anti-BE445 antibodies in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer is provided.

30           The BARD1, B123, BE2, BE14, BE31 or BE445 genes and proteins of the present invention have many utilities. For example, their BRCA1 binding properties may be exploited

in methods to detect BRCA1 proteins. Such methods comprise contacting a sample suspected of containing a BRCA1 protein with a BRCA1-binding BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide or fusion protein, under conditions effective to allow the formation of BRCA1-BARD1, -B123, -BE2, -BE14, -BE31 or -BE445 complexes, and detecting the  
5 BRCA1-BARD1, -B123, -BE2, -BE14, -BE31 or -BE445 complexes so formed.

Methods of purifying BRCA1 proteins are also provided, which comprise contacting a composition comprising a BRCA1 protein with a BRCA1-binding BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide or fusion protein, under conditions effective to allow the  
10 formation of BRCA1-BARD1, -B123, -BE2, -BE14, -BE31 or -BE445 complexes, and obtaining the BRCA1 protein from the BRCA1-BARD1, -B123, -BE2, -BE14, -BE31 or -BE445 complexes in a more purified form.

The "BRCA1-binding BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide or  
15 fusion proteins" of such methods are any BARD1, B123, BE2, BE14, BE31 or BE445 proteins or fragments sufficient to operatively bind BRCA1, using the assays and criteria disclosed herein.

Certain methods for detecting BARD1, B123, BE2, BE14, BE31 or BE445 in a sample  
20 comprise contacting a sample suspected of containing BARD1, B123, BE2, BE14, BE31 or BE445 with a first antibody that binds to a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide, or a mutant thereof, under conditions effective to allow the formation of immune complexes, and detecting the immune complexes thus formed. In addition to their diagnostic use, these methods are also suitable for purifying BARD1, B123, BE2, BE14, BE31 or BE445,  
25 identifying BARD1, B123, BE2, BE14, BE31 or BE445 expression, in identifying engineered mutants and in titrating BARD1, B123, BE2, BE14, BE31 or BE445 and/or BARD1, B123, BE2, BE14, BE31 or BE445 antibodies.

The invention further provides diagnostic methods, particularly useful in connection with  
30 breast, ovarian and uterine cancer, but also of potential usefulness in other cancers, particularly lung, colon and other cancers.

Diagnostically, the present invention provides methods for identifying a patient having or at risk for developing breast, ovarian or uterine cancer, comprising determining the type or amount of BARD1, B123, BE2, BE14, BE31 or BE445 present within a biological sample from the patient, wherein the presence of a BARD1, B123, BE2, BE14, BE31 or BE445 mutant or an altered amount of wild-type BARD1, B123, BE2, BE14, BE31 or BE445, in comparison to a sample from a normal subject, is indicative of a patient having or at risk for developing breast, ovarian or uterine cancer.

The "type" of BARD1, B123, BE2, BE14, BE31 or BE445 may be determined, allowing mutant genes and proteins to be distinguished from wild-types. The use of mutant- and wild-type-specific nucleic acid probes is particularly contemplated. In the beginning, the use of wild-type-specific nucleic acid probes will be preferred. The identification of a particularly diagnostic mutant sequence will then lead to the increased use of that mutant sequence, either in the population or in defined families. The use of mutant- and wild-type-specific antibodies is also contemplated, as may be prepared using mutant- and wild-type-specific BARD1, B123, BE2, BE14, BE31 or BE445 peptides.

Where the "amount" of BARD1, B123, BE2, BE14, BE31 or BE445 is determined, a lesser amount of the natural BARD1, B123, BE2, BE14, BE31 or BE445 protein may be indicative of the propensity to develop breast, ovarian or uterine cancer, as is typical with tumor suppressors. A greater amount of BARD1, B123, BE2, BE14, BE31 or BE445 could also be indicative of the propensity to develop breast, ovarian or uterine cancer, which situation would represent the case where the BARD1, B123, BE2, BE14, BE31 or BE445 is a dominant proto-oncogene. In any event, changes from the naturally observed range in the population will be easily detected and will have implications for disease risk and development.

The type or amount of BARD1, B123, BE2, BE14, BE31 or BE445 may be determined by means of a molecular biological assay to determine the type or amount of a nucleic acid that encodes BARD1, B123, BE2, BE14, BE31 or BE445. Such molecular biological assays will often comprise a direct or indirect step that allows a determination of the sequence of at least a portion of the BARD1-, B123-, BE2-, BE14-, BE31- or BE445-encoding nucleic acid, which sequence can be compared to a wild-type BARD1, B123, BE2, BE14, BE31 or BE445 sequence,

such as SEQ ID NO:1, SEQ ID NO:17, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46 or another acceptable normal allelic or polymorphic sequence, such as, in the case of BARD1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38.

5

It is contemplated that BARD1, B123, BE2, BE14, BE31 or BE445 sequences diagnostic or prognostic for breast, ovarian, uterine or even for other forms of cancer may comprise at least one point mutation, deletion, translocation, insertion, duplication or other aberrant change. Diagnostic RFLPs are thus also contemplated. RNase protection assays may also be employed in certain embodiments.

10

Diagnostic methods may be based upon the steps of:

(a) obtaining a biopsy sample from a subject or patient;

15

(b) contacting sample nucleic acids from the biopsy sample with an isolated BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segment under conditions effective to allow hybridization of substantially complementary nucleic acids; and

20

(c) detecting, and optionally further characterizing, the hybridized complementary nucleic acids thus formed.

The methods may involve *in situ* detection of sample nucleic acids located within the cells of the sample. The sample nucleic acids may also be separated from the cell prior to contact. The sample nucleic acids may be DNA or RNA.

25

The methods may involve the use of isolated BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segments that comprises a radio, enzymatic or fluorescent detectable label, wherein the hybridized complementary nucleic acids are detected by detecting the label.

30

PCR® will often be preferred, as exemplified by the steps of:

- 5 (a) contacting the sample nucleic acids with a pair of nucleic acid primers that hybridize to distant sequences from a mutant, polymorphic or wild-type BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, the primers capable of amplifying a mutant, polymorphic or wild-type BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid segment when used in conjunction with a polymerase chain reaction;
- 10 (b) conducting a polymerase chain reaction to create amplification products; and
- (c) detecting and characterizing the amplification products thus formed.

15 Diagnostic immunoassay methods are also provided, wherein the type or amount of BARD1, B123, BE2, BE14, BE31 or BE445 is determined by means of an immunoassay to determine the type or amount of a BARD1, B123, BE2, BE14, BE31 or BE445 protein. Such methods may comprise the steps of:

- 20 (a) obtaining a biopsy sample from a subject or patient;
- (b) contacting the biopsy sample with a first antibody that binds to a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide, or mutant, under conditions effective to allow the formation of specific immune complexes; and
- 25 (c) detecting the specific immune complexes thus formed.

30 The first antibody may be linked to a detectable label, wherein the immune complexes are directly detected by detecting the presence of the label. The immune complexes may also be indirectly detected by means of a second antibody linked to a detectable label, the second antibody having binding affinity for the first antibody.

Where BARD1, B123, BE2, BE14, BE31 or BE445 proves to be a tumor suppressor, the present invention also provides methods of treating cancers such as breast, ovarian or uterine

cancer, comprising administering to a patient with breast, ovarian or uterine cancer a biologically effective amount of a pharmaceutically acceptable BARD1, B123, BE2, BE14, BE31 or BE445 composition

5           Where BARD1, B123, BE2, BE14, BE31 or BE445 proves to be an oncogene, the invention further provides methods of treating cancers such as breast, ovarian or uterine cancer, comprising administering to a patient with breast, ovarian or uterine cancer a biologically effective amount of a pharmaceutically acceptable composition that inhibits BARD1, B123, BE2, BE14, BE31 or BE445. The composition may comprises a component that inhibits a  
10 BARD1, B123, BE2, BE14, BE31 or BE445 gene, mRNA, protein, peptide or BRCA1-BARD1, -B123, -BE2, -BE14, -BE31 or -BE445 complex. Examples of inhibitors include antisense constructs, ribozymes, inhibitory antibodies, and recombinant vectors that express any of the foregoing BARD1, B123, BE2, BE14, BE31 or BE445 inhibitors in mammalian cells.

15           The tumor suppressor-type treatment may also comprise giving BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide compositions or BARD1, B123, BE2, BE14, BE31 or BE445 DNA segments or recombinant vectors that expresses BARD1, B123, BE2, BE14, BE31 or BE445 proteins or peptides in the target cells. Enhancing BARD1, B123, BE2, BE14, BE31 or BE445 transcription, translation or stability is also contemplated.

20           The cancer treatment methods of the present invention may be combined with any standard anti-cancer strategy, such as surgery, chemotherapy, radiotherapy and other gene therapies. The administration of a biologically effective amount of a BRCA1 protein, peptide or recombinant vector composition is also contemplated.

25           The present invention also provides BARD1, B123, BE2, BE14, BE31 and BE445 nucleic acid segments, proteins, polypeptides, peptides, domains and fusion proteins for use in the preparation of a prophylactic formulation for administration to a patient at risk for developing cancer or a patient in the early stages of cancer. Thus, the use of BARD1, B123,  
30 BE2, BE14, BE31 and BE445 nucleic acid segments, proteins, polypeptides, peptides, domains and fusion proteins in the preparation of a prophylactic formulation for administration to a patient at risk for developing cancer or a patient in the early stages of cancer is provided.

Additionally, the present invention provides a nucleic acid segment for use in the preparation of a medicament for use in treating a patient with cancer. Therefore, the use of a nucleic acid segment in the preparation of a medicament for use in treating a patient with cancer is also provided.

5

In that the BARD1, B123, BE2, BE14, BE31 or BE445 and BRCA1 interaction is important for BRCA1 and BARD1, B123, BE2, BE14, BE31 or BE445 function, the present invention further provides methods for identifying a BARD1, B123, BE2, BE14, BE31, BE445 or BRCA1 agonist or stimulant, or antagonist or inhibitor, comprising contacting a composition comprising BARD1, B123, BE2, BE14, BE31 or BE445 and BRCA1 with a candidate substance and identifying a candidate substance that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 and BRCA1 or that alters the activity, such as the DNA binding, transcriptional or other functional activity, of a BARD1-, B123-, BE2-, BE14-, BE31- or BE445-BRCA1 bound complex. The BARD1, B123, BE2, BE14, BE31 or BE445 or BRCA1 agonists or antagonists prepared by such as process form another aspect of the present invention, which substances may also be employed in treating breast, ovarian or uterine cancer.

Thus, the present invention also provides BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains and fusion proteins for use in the identification of a binding protein agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or that alters biological activity of a BRCA1-BARD1, BRCA1-B123, BRCA1-BE2, BRCA1-BE14, BRCA1-BE31 or BRCA1-BE445 complex. Therefore, the use of BARD1, B123, BE2, BE14, BE31 and BE445 proteins, polypeptides, peptides, domains and fusion proteins in the identification of a binding protein agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or that alters biological activity of a BRCA1-BARD1, BRCA1-B123, BRCA1-BE2, BRCA1-BE14, BRCA1-BE31 or BRCA1-BE445 complex is provided.



### **BRIEF DESCRIPTION OF THE DRAWINGS**

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by  
5 reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

**FIG. 1.** Mammalian two-hybrid analysis of interaction between BR304 and the candidate BRCA1-associated polypeptides. Each culture of 293 cells was transiently co-transfected with the G5LUC reporter plasmid and the two indicated expression vectors. The  
10 GAL4 expression vector encoded either the "parental" GAL4 DNA-binding domain (denoted by "+" in the GAL4 column) or the GAL4-BR304 hybrid polypeptide. The VP16 expression vector encoded either the parental VP16 transactivation domain (denoted by "+" in the VP16 column) or the indicated VP16-hybrid polypeptide. Duplicate transfections were conducted for each  
15 combination of expression plasmids, and the normalized luciferase activities obtained from each transfection are illustrated.

**FIG. 2.** A schematic comparison of the BRCA1 and BARD1 polypeptides. The map of BRCA1 illustrates sequences that comprise the RING motif (20-68) and the BRCT domain (1685-1863); the N-terminal and C-terminal core motifs of the BRCT domain (residues 1699-  
20 1736 and 1818-1855, respectively) are denoted by the solid bars marked "n" and "c", respectively. The map of the BARD1 illustrates the RING motif (residues 44-90), the three ankyrin repeats (residues 427-525), and the BRCT domain (residues 605-777); the N-terminal and C-terminal core motifs of the BRCT domain (residues 616-653 and 743-777, respectively)  
25 are denoted by the solid bars marked "n" and "c", respectively. The sequences encoded by the B202 and B230 cDNA clones are indicated beneath the BARD1 map. The NE (residues 26-142) and NB (residues 26-202) segments of BARD1 used in FIG. 3 are also shown.

**FIG. 3.** Mammalian two-hybrid analysis of the interaction between BRCA1 and defined  
30 segments of the BARD1 polypeptide. Each dish of 293 cells was transiently co-transfected with the G5LUC reporter plasmid, the pSV- $\beta$ -galactosidase control plasmid, and the two indicated

expression vectors. The GAL4 expression vector encoded either the "parental" GAL4 DNA-binding domain (denoted by "+" in the GAL4 column) or the GAL4-BR304 hybrid polypeptide. The VP16 expression vector encoded either the parental VP16 transactivation domain (denoted by "+" in the VP16 column) or the VP16-hybrid polypeptide containing segments NE (residues 26-142) or NB (residues 26-202) of BARD1 (see FIG. 2).

**FIG. 4A and FIG. 4B.** BRCA1 sequences that mediate association with BARD1. **FIG. 4A**, mammalian two-hybrid analysis of the interaction between BARD1 and defined segments of BRCA1. Each dish of 293 cells was transiently co-transfected with the G5LUC reporter plasmid, the pSV- $\beta$ -galactosidase control plasmid, and the two indicated expression vectors. The VP16 expression vector encoded either the "parental" VP16 transactivation domain (denoted by "+" in the VP16 column) or VP16-NE, a hybrid polypeptide containing amino acids 26-142 of BARD1. The GAL4 expression vector encoded either the parental GAL4 DNA-binding domain (denoted by "+" in the GAL4 column) or the indicated GAL4-hybrid polypeptide; the latter contained BRCA1 residues 1-147 (BR147), 1-101 (BR101), 1-71 (BR71), or 1-45 (BR45). **FIG. 4B**, a reciprocal two-hybrid analysis of BARD1 interaction with defined segments of BRCA1. The GAL4 expression vector encoded either the parental GAL4 DNA-binding domain (denoted by "+" in the GAL4 column) or GAL4-NE, a hybrid polypeptide containing amino acids 26-142 of BARD1. The VP16 expression vector encoded either the parental VP16 transactivation domain (denoted by "+" in the VP16 column) or a VP16-hybrid polypeptide containing the indicated segment of BRCA1.

**FIG. 5A and FIG. 5B.** Tumorigenic mutants of BRCA1 fail to interact with BARD1. **FIG. 5A**, mammalian two-hybrid analysis of the interaction between BARD1 and the mutant derivatives of BRCA1. Each dish of 293 cells was transiently co-transfected with the G5LUC reporter plasmid, the pSV- $\beta$ -galactosidase control plasmid, and the two indicated expression vectors. The VP16 expression vector encoded either the parental VP16 transactivation domain (denoted by "+" in the VP16 column) or VP16-NE, a hybrid polypeptide containing amino acids 26-142 of BARD1. The GAL4 expression vector encoded either the "parental" GAL4 DNA-binding domain (denoted by "+" in the GAL4 column) or the indicated GAL4-BR304 fusion protein; the latter included wild-type BRCA1 residues 1-304 (BR304; lanes 3 and 4) and variants of BR304 that bear the tumorigenic C61G or C64G mutations (lanes 5-8). **FIG. 5B**, co-

immunoprecipitation analysis of the interaction between BARD1 and the mutant derivatives of BRCA1. 293 cells were transfected with a pair of expression vectors encoding FLAG-B202 and either a wild-type or mutant derivative of FLAG-BR304. After two days the cells were lysed and the lysates were normalized for expression of FLAG-B202. Equivalent aliquots of the lysates (100 ml) were immunoprecipitated with the BRCA1-specific antiserum (lanes 2, 4, and 6) or the corresponding pre-immune serum (lanes 1, 3, and 5). The immunoprecipitates were then fractionated by SDS-PAGE, and the FLAG-B202 and FLAG-BR304 polypeptides were detected by immunoblotting with the M5 monoclonal antibody. As shown, FLAG-B202 was co-immunoprecipitated with the wild-type FLAG-BR304 (lane 2) but not with derivatives of FLAG-BR304 containing the C61G (lane 4) or C64G (lane 6) mutation. Expression of the different FLAG-BR304 derivatives was compared by immunoblotting equivalent aliquots (20 ml) of the untreated lysates with FLAG-specific M5 monoclonal antibody (Eastman Kodak) (lanes 7-9).

**FIG. 6.** Schematic diagram of the BARD1 cDNA. The ring domain, ankyrin repeats, BRCT domain and 5' and 3' untranslated regions are shaded as indicated. Splice sites are designated A-H. The location of the splice site according to the nucleotide sequence of the gene (GenBank Accession No. U76638) or the amino acid sequence of the protein are indicated above the diagram. Additional splice sites exist between G and H but these have not yet been determined. Mutations described in this manuscript are indicated above the cDNA diagram. Polymorphisms are indicated below the diagram. Designations of amino acid changes are according to the nomenclature proposed by Beaudet and Tsui (1993).

### SEQUENCE SUMMARY

SEQ ID NO:1	BARD1 DNA Sequence
SEQ ID NO:2	BARD1 Amino Acid Sequence
SEQ ID NO:3	FLAG Epitope Amino Acid Sequence
SEQ ID NO:4	5' Primer for PCR Amplification of N-terminus of BRCA1
SEQ ID NO:5	3' Primer for PCR Amplification of N-terminus of BRCA1
SEQ ID NO:6	BARD1 PCR Primer B202L
SEQ ID NO:7	BARD1 PCR Primer B202R

	SEQ ID NO:8	HA-BR304 Amino Terminal Tag Amino Acid Sequence
	SEQ ID NO:9	TCL52 DNA Sequence
	SEQ ID NO:10	TCL163 DNA Sequence
	SEQ ID NO:11	B223 DNA Sequence
5	SEQ ID NO:12	B115 DNA Sequence
	SEQ ID NO:13	BAP28 DNA Sequence
	SEQ ID NO:14	B48 DNA Sequence
	SEQ ID NO:15	B258 DNA Sequence
	SEQ ID NO:16	BAP152 DNA Sequence
10	SEQ ID NO:17	B123 DNA Sequence
	SEQ ID NO:18	B268 DNA Sequence
	SEQ ID NO:19	B123 Amino Acid Sequence
	SEQ ID NO:20	BARD1 P143 DNA Sequence
	SEQ ID NO:21	BARD1 P143 Amino Acid Sequence
15	SEQ ID NO:22	BARD1 P553 DNA Sequence
	SEQ ID NO:23	BARD1 P553 Amino Acid Sequence
	SEQ ID NO:24	BARD1 P1121 DNA Sequence
	SEQ ID NO:25	BARD1 P1121 Amino Acid Sequence
	SEQ ID NO:26	BARD1 PΔ1140-1160 DNA Sequence
20	SEQ ID NO:27	BARD1 PΔ1140-1160 Amino Acid Sequence
	SEQ ID NO:28	BARD1 P1592 DNA Sequence
	SEQ ID NO:29	BARD1 P1592 Amino Acid Sequence
	SEQ ID NO:30	BARD1 P1765 DNA Sequence
	SEQ ID NO:31	BARD1 P1765 Amino Acid Sequence
25	SEQ ID NO:32	BARD1 MQ564H DNA Sequence
	SEQ ID NO:33	BARD1 MQ564H Amino Acid Sequence
	SEQ ID NO:34	BARD1 MS761N DNA Sequence
	SEQ ID NO:35	BARD1 MS761N Amino Acid Sequence
	SEQ ID NO:36	BARD1 MR658C DNA Sequence
30	SEQ ID NO:37	BARD1 MR658C Amino Acid Sequence
	SEQ ID NO:38	BARD1 P2354 DNA Sequence
	SEQ ID NO:39	BARD1 P2354 Amino Acid Sequence

	SEQ ID NO:40	BE2 DNA Sequence
	SEQ ID NO:41	BE2 Amino Acid Sequence
	SEQ ID NO:42	BE14 DNA Sequence
	SEQ ID NO:43	BE14 Amino Acid Sequence
5	SEQ ID NO:44	BE31 DNA Sequence
	SEQ ID NO:45	BE31 Amino Acid Sequence
	SEQ ID NO:46	BE445 DNA Sequence
	SEQ ID NO:47	BE445 Amino Acid Sequence
	SEQ ID NO:48	TCL52 Amino Acid Sequence
10	SEQ ID NO:49	TCL163 Amino Acid Sequence
	SEQ ID NO:50	B223 Amino Acid Sequence
	SEQ ID NO:51	B115 Amino Acid Sequence
	SEQ ID NO:52	BAP28 Amino Acid Sequence
	SEQ ID NO:53	B48 Amino Acid Sequence
15	SEQ ID NO:54	B258 Amino Acid Sequence
	SEQ ID NO:55	BAP152 Amino Acid Sequence
	SEQ ID NO:56	B268 Amino Acid Sequence
	SEQ ID NO:57	BARD1 PCR Primer R135S
	SEQ ID NO:58	BARD1 PCR Primer R135AS
20	SEQ ID NO:59	BARD1 PCR Primer B202-Z1S
	SEQ ID NO:60	BARD1 PCR Primer B202-ZAS
	SEQ ID NO:61	BARD1 PCR Primer B202-Z1SP
	SEQ ID NO:62	BARD1 PCR Primer B202-A
	SEQ ID NO:63	BARD1 PCR Primer B202-N
25	SEQ ID NO:64	BARD1 PCR Primer B202-B
	SEQ ID NO:65	BARD1 PCR Primer B202-BAS
	SEQ ID NO:66	BARD1 PCR Primer B202-X
	SEQ ID NO:67	BARD1 PCR Primer B202-XAS
	SEQ ID NO:68	BARD1 PCR Primer B230-A
30	SEQ ID NO:69	BARD1 PCR Primer B230-AS
	SEQ ID NO:70	BARD1 PCR Primer B202-Y
	SEQ ID NO:71	BARD1 PCR Primer B202-YAS

	SEQ ID NO:72	BARD1 PCR Primer B230-B
	SEQ ID NO:73	BARD1 PCR Primer B230-BAS
	SEQ ID NO:74	BARD1 PCR Primer B230-C
	SEQ ID NO:75	BARD1 PCR Primer B230-CAS
5	SEQ ID NO:76	BARD1 PCR Primer B230-D
	SEQ ID NO:77	BARD1 PCR Primer B230-DAS
	SEQ ID NO:78	BARD1 PCR Primer B230-PS
	SEQ ID NO:79	BARD1 PCR Primer B230-P
	SEQ ID NO:80	BARD1 PCR Primer B230-E
10	SEQ ID NO:81	BARD1 PCR Primer B230-EAS
	SEQ ID NO:82	BARD1 PCR Primer B230-F
	SEQ ID NO:83	BARD1 PCR Primer B230-FAS
	SEQ ID NO:84	BARD1 PCR Primer B230-FF
	SEQ ID NO:85	BARD1 PCR Primer B230-FFAS
15	SEQ ID NO:86	BARD1 PCR Primer B230-WS
	SEQ ID NO:87	BARD1 PCR Primer B230-WAS
	SEQ ID NO:88	BARD1 PCR Primer B230-G
	SEQ ID NO:89	BARD1 PCR Primer B230-H
	SEQ ID NO:90	BARD1 PCR Primer B230-HAS
20	SEQ ID NO:91	BARD1 PCR Primer B230-TS
	SEQ ID NO:92	BARD1 PCR Primer B230-TAS
	SEQ ID NO:93	BARD1 PCR Primer B230-US
	SEQ ID NO:94	BARD1 PCR Primer B230-UAS
	SEQ ID NO:95	BARD1 PCR Primer R1352S
25	SEQ ID NO:96	BARD1 PCR Primer R13AAS
	SEQ ID NO:97	BARD1 PCR Primer R12AS
	SEQ ID NO:98	BARD1 PCR Primer R12BAS
	SEQ ID NO:99	BARD1 PCR Primer R13B5
	SEQ ID NO:100	BARD1 PCR Primer R13CAS
30	SEQ ID NO:101	BARD1 PCR Primer R5C5
	SEQ ID NO:102	BARD1 PCR Primer B202-N1
	SEQ ID NO:103	BARD1 PCR Primer R5DAS

	SEQ ID NO:104	BARD1 PCR Primer R34D5
	SEQ ID NO:105	BARD1 PCR Primer R34FAS
	SEQ ID NO:106	BARD1 PCR Primer R34F5
	SEQ ID NO:107	BARD1 PCR Primer R34GA5
5	SEQ ID NO:108	BARD1 PCR Primer R36H5
	SEQ ID NO:109	BARD1 PCR Primer R36EAS
	SEQ ID NO:110	BARD1 PCR Primer R36E5
	SEQ ID NO:111	BAP152 Second Amino Acid Sequence
	SEQ ID NO:112	BRCA1 PCR Primer 4L
10	SEQ ID NO:113	BRCA1 PCR Primer 4R
	SEQ ID NO:114	BRCA2 Forward PCR Primer
	SEQ ID NO:115	BRCA2 Reverse PCR Primer
	SEQ ID NO:116	BARD1 PCR Primer FFGS2
	SEQ ID NO:117	BARD1 PCR Primer B2305FGAS
15	SEQ ID NO:118	BARD1 PCR Primer 3FGR
	SEQ ID NO:119	BARD1 PCR Primer WSGAS
	SEQ ID NO:120	BARD1 PCR Primer B230IXS
	SEQ ID NO:121	BARD1 PCR Primer B230IXAS
	SEQ ID NO:122	BARD1 Genomic DNA Contig 1 (Contains Exon 1)
20	SEQ ID NO:123	BARD1 Genomic DNA Contig 2 (Contains Exon 2 and Exon3)
	SEQ ID NO:124	BARD1 Genomic DNA Contig 3 (Contains Exon 4)
	SEQ ID NO:125	BARD1 Genomic DNA Contig 4 (Contains Exon 5)
	SEQ ID NO:126	BARD1 Genomic DNA Contig 5 (Contains Exon 6)
	SEQ ID NO:127	BARD1 Genomic DNA Contig 6 (Contains Exon 7)
25	SEQ ID NO:128	BARD1 Genomic DNA Contig 7 (Contains Exon 8)
	SEQ ID NO:129	BARD1 Genomic DNA Contig 8 (Contains Exon 9)
	SEQ ID NO:130	BARD1 Genomic DNA Contig 9 (Contains Exon 10 and Exon 11)

## DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

In order to identify proteins that bind to BRCA1, the inventors first utilized the yeast two-hybrid system to identify proteins that associate with BRCA1 *in vivo* (Fields and Song, 1989; Chien *et al.*, 1991; Durfee *et al.*, 1993; Harper *et al.*, 1993). Such analyses led to the discovery of fifteen novel genes that encode polypeptides that bind to the N-terminal 304 amino acids of BRCA1 in the yeast assay.

These are included herein as BARD1 DNA and protein sequences (SEQ ID NO:1 and SEQ ID NO:2, respectively); and also TCL52 DNA sequence (SEQ ID NO:9); TCL163 DNA sequence (SEQ ID NO:10); B223 DNA sequence (SEQ ID NO:11); B115 DNA sequence (SEQ ID NO:12); BAP28 DNA sequence (SEQ ID NO:13); B48 DNA sequence (SEQ ID NO:14); B258 DNA sequence (SEQ ID NO:15); BAP152 DNA sequence (SEQ ID NO:16); B123 DNA and protein sequences (SEQ ID NO:17 and SEQ ID NO:19, respectively); B268 DNA sequence (SEQ ID NO:18); BE2 DNA and protein sequences (SEQ ID NO:40 and SEQ ID NO:41, respectively); BE14 DNA and protein sequences (SEQ ID NO:42 and SEQ ID NO:43, respectively); BE31 DNA and protein sequences (SEQ ID NO:44 and SEQ ID NO:45, respectively); and BE445 DNA and protein sequences (SEQ ID NO:46 and SEQ ID NO:47, respectively). Each of the genes and proteins listed above are included within all aspects of the present invention.

The yeast screening assay also led to the identification of five further gene and protein candidates for BRCA1 binding. Although the sequences of these five genes have been previously reported, their potential role in BRCA1 binding and/or as part of the breast cancer development pathway(s) has not previously been suggested. As such, the genes and proteins TAFII70/80 (Genbank accession nos. L25444 and U31659), filamin (X53416), STAT3/APRF (L29277), UNPH (U20657), and a human homolog of the yeast GCN5 gene product (U57317), are each included within the methodological aspects of the present invention to the extent that such methods could not previously have been contemplated.



To even further increase the chances that the yeast screening assay resulted in the identification of protein interactions that are physiologically-relevant, rather than just artifactual results of over-expression of foreign proteins in yeast, the inventors used a mammalian two-hybrid assay (Dang *et al.*, 1991). The mammalian assay appears to be especially stringent; thus, although false-negative results were observed in previous studies with this method, false-positive results have not as yet been reported (Altschul *et al.*, 1990).

Of the fifteen analyzed candidate BRCA1-associated proteins identified by two-hybrid screening in yeast (11 novel and 5 known sequences), each protein tested except that encoded by a clone termed B202 failed to associate with BRCA1 in the mammalian assay (the sixteenth candidate, laminin, has not yet been tested). A second independent isolate (B230) was also obtained that contained a distinct but overlapping insert of 2.5 kb. The combined B202 and B230 cDNA sequence of 2,531 bp (SEQ ID NO:1) was termed the *BARD1* gene, and this gene encodes the 777 and/or 752 amino acid protein of SEQ ID NO:2, also termed BARD1 (named from BRCA1-Associated RING Domain (BARD1) protein, see below).

As only BARD1 registered as positive in the mammalian assay, this gene and protein are naturally the currently preferred biological compositions for use in the present invention. However, as false-negative results have been encountered previously in mammalian two-hybrid studies, the inability of the other fourteen (or fifteen) proteins to interact with BRCA1 in this assay does not necessarily eliminate them as candidate BRCA1-associated factors. It is for this reason that they are still encompassed within all aspects of the present invention. Even though one or more, or even nearly all, of the additionally disclosed proteins may ultimately prove not to bind to BRCA1, this would not negate the usefulness of one or two proteins or more from the remaining candidates upon confirmation of BRCA1-binding properties for such proteins.

In any event, the interaction between BRCA1 and BARD1 was detected in both orientations of the mammalian two-hybrid system, and it was confirmed in an independent fashion by co-immunoprecipitation of these proteins from mammalian cell lysates. Furthermore, the *in vivo* association between these proteins was reproduced using *in vitro* assays of protein binding, indicating that the interaction between BRCA1 and BARD1 is direct. Therefore, the utility of BARD1 in BRCA1 binding has been rigorously shown.

The BARD1 protein is a novel RING protein that interacts with the amino-terminal region of BRCA1. The BRCA1-associated RING domain (BARD1) protein is encoded by sequences on chromosome 2q, and resembles BRCA1 in that it possesses an amino-terminal RING motif and the carboxy-terminal BRCT domains. These features, as well as its ability to form *in vivo* complexes with BRCA1, indicate that BARD1 gene and protein likely serves as an effector and/or a regulator of BRCA1-mediated tumor suppression.

The precise role of BARD1 in tumor formation is not yet known, although this does not negate the usefulness of the BARD1 compositions of the present invention, particularly and most immediately, in terms of diagnostics. On one hand, tumor suppression may be mediated by the protein complex formed by the interaction between BRCA1 and BARD1. As such, BARD1 would itself function as a tumor suppressor.

The tumor suppressor model is appealing because many regulatory proteins are known to function as obligate heterodimers, including transcription factors implicated in cancer, such as the c-MYC protein (which functions as a transcription factor within the context of a c-MYC/MAX heterodimer). If BARD1 is confirmed to be tumor suppressor, the provision of wild-type BARD1 to a cancer cell should counteract the malignant phenotype. As such, breast cancer treatment would include administering BARD1 to a patient.

On the other hand, it is well established that certain dominant proto-oncogenes promote tumorigenesis by binding and reducing the activity of tumor suppressor proteins. Prominent examples include MDM2, which binds and inhibits the tumor suppressor function of p53, and the transforming proteins encoded by certain DNA viruses (*e.g.*, the SV40 large T antigen), that also bind and inactivate tumor suppressors such as p53 and Rb. Thus, it is formally possible that the interaction between BARD1 and BRCA1 would reduce the tumor suppressor function of BRCA1.

In the above scenario, the gene encoding BARD1 would serve as a dominant proto-oncogene. If BARD1 is confirmed to be a classical oncogene, inhibiting BARD1 would be the

therapeutic approach. BARD1 inhibition could be achieved by providing to a cancer cell or administering to a patient any compound that inhibits the BARD1 gene, mRNA or protein.

5 The diagnostic and therapeutic methods disclosed herein take account of both the candidate tumor suppressor and oncogenic properties of BARD1 and the other BRCA1 binding proteins of the present invention.

#### **I. BARD1 and Other BRCA1 Binding Proteins: Genes and DNA Segments**

10 Important aspects of the present invention concern isolated DNA segments and recombinant vectors encoding wild-type, polymorphic or mutant BARD1, and the creation and use of recombinant host cells through the application of DNA technology, that express wild-type, polymorphic or mutant BARD1, using sequences of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID  
15 NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. DNA segments, recombinant vectors, recombinant host cells and expression methods involving the other BRCA1 binding proteins of the present invention, using sequences of TCL52 (SEQ ID NO:9); TCL163 (SEQ ID NO:10); B223 (SEQ ID NO:11);  
20 B115 (SEQ ID NO:12); BAP28 (SEQ ID NO:13); B48 (SEQ ID NO:14); B258 (SEQ ID NO:15); BAP152 (SEQ ID NO:16); B123 (SEQ ID NO:17); B268 (SEQ ID NO:18); BE2 (SEQ ID NO:40); BE14 (SEQ ID NO:42); BE31 (SEQ ID NO:44); and BE445 (SEQ ID NO:46) are also provided. Each of the foregoing genes are included within all aspects of the following description.

25

The present invention concerns DNA segments, isolatable from mammalian and human cells, that are free from total genomic DNA and that are capable of expressing a protein or polypeptide that has BRCA1-binding activity.

30

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding BARD1 refers to a DNA segment that contains wild-type, polymorphic or mutant BARD1,

TCL52, TCL163, B223, B115, BAP28, B48, B258, BAP152, B123, B268, BE2, BE14, BE31 or BE445 coding sequences yet is isolated away from, or purified free from, total mammalian or human genomic DNA. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified wild-type, polymorphic or mutant BARD1 or BRCA1-binding protein gene refers to a DNA segment including wild-type, polymorphic or mutant BARD1 or BRCA1-binding protein coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins and mutants.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case the wild-type, polymorphic or mutant BARD1 gene, or other BRCA1 binding protein genes, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode a wild-type, polymorphic or mutant BARD1 protein or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially as set forth in, SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, corresponding to wild-type, polymorphic or mutant human BARD1. Moreover, in other particular embodiments, the invention concerns isolated DNA segments and recombinant vectors that encode a BARD1

protein or peptide that includes within its amino acid sequence the substantially full length protein sequence of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39.

5

In other embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode a BRCA1 binding protein or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially as set forth in, any one of SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47, corresponding to the human BRCA1 binding proteins TCL52, TCL163, B223, B115, BAP28, B48, B258, BAP152, B123, B268, BE2, BE14, BE31 or BE445. Moreover, in other particular embodiments, the invention concerns isolated DNA segments and recombinant vectors that encode a BRCA1 binding protein or peptide that includes within its amino acid sequence the substantially full length protein sequence of SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47.

The term "a sequence essentially as set forth in SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47" means that the sequence substantially corresponds to a portion of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47.

The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein. Accordingly, sequences that have between about 70% and about 80%; or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47 will be sequences that are "essentially as set forth in SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:48 through SEQ ID NO:56, SEQ ID NO:19, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45 or SEQ ID NO:47", provided the biological activity of the protein is maintained.

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. The term "essentially as set forth in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130 and has relatively few codons that are not

identical, or functionally equivalent, to the codons of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID  
5 NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. Again, DNA segments that encode proteins exhibiting BRCA1-binding activity will be most preferred.

The term "functionally equivalent codon" is used herein to refer to codons that encode  
10 the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (see Table 1, below).

**Table 1 - Preferred Human DNA Codons**

<u>Amino Acids</u>			<u>Codons</u>						
Alanine	Ala	A	GCC	GCT	GCA	GCG			
Cysteine	Cys	C	TGC	TGT					
Aspartic acid	Asp	D	GAC	GAT					
Glutamic acid	Glu	E	GAG	GAA					
Phenylalanine	Phe	F	TTC	TTT					
Glycine	Gly	G	GGC	GGG	GGA	GGT			
Histidine	His	H	CAC	CAT					
Isoleucine	Ile	I	ATC	ATT	ATA				
Lysine	Lys	K	AAG	AAA					
Leucine	Leu	L	CTG	CTC	TTG	CTT	CTA	TTA	
Methionine	Met	M	ATG						
Asparagine	Asn	N	AAC	AAT					
Proline	Pro	P	CCC	CCT	CCA	CCG			
Glutamine	Gln	Q	CAG	CAA					
Arginine	Arg	R	CGC	AGG	CGG	AGA	CGA	CGT	
Serine	Ser	S	AGC	TCC	TCT	AGT	TCA	TCG	
Threonine	Thr	T	ACC	ACA	ACT	ACG			
Valine	Val	V	GTG	GTC	GTT	GTA			
Tryptophan	Trp	W	TGG						
Tyrosine	Tyr	Y	TAC	TAT					

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.



Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences that have between about 70% and about 79%; or more preferably, between about 80% and about 89%; or even more preferably, between about 90% and about 99%; of nucleotides that are identical to the nucleotides of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130 will be sequences that are "essentially as set forth in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130".

Sequences that are essentially the same as those set forth in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art, as disclosed herein..

Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1, any

one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. Nucleic acid sequences that are "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130 under relatively stringent conditions such as those described herein.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

For example, nucleic acid fragments may be prepared that include a short contiguous stretch identical to or complementary to SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, such as about 8, about 10 to about 14, or about 15 to about 20 nucleotides, and that are up to about 20,000, or about 10,000, or about 5,000 base pairs in

length, with segments of about 3,000 being preferred in certain cases. DNA segments with total lengths of about 1,000, about 500, about 200, about 100 and about 50 base pairs in length (including all intermediate lengths) are also contemplated to be useful.

5 It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etc.; 21, 22, 23, etc.; 30, 31, 32, etc.; 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; including all integers through the 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000; 5,000-10,000 ranges, up to and including sequences of about 12,001, 12,002, 13,001, 13,002,  
10 15,000, 20,000 and the like.

The various probes and primers designed around the disclosed nucleotide sequences of the present invention may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers can be  
15 proposed:

$$n \text{ to } n + y$$

where  $n$  is an integer from 1 to the last number of the sequence and  $y$  is the length of the primer  
20 minus one, where  $n + y$  does not exceed the last number of the sequence. Thus, for a 10-mer, the probes correspond to bases 1 to 10, 2 to 11, 3 to 12 ... and so on. For a 15-mer, the probes correspond to bases 1 to 15, 2 to 16, 3 to 17 ... and so on. For a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 ... and so on.

25 It will also be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID  
30 NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. Recombinant vectors and isolated DNA segments may therefore variously include these coding regions themselves, coding regions bearing selected alterations or

modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include such coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

5           The DNA segments of the present invention encompass biologically functional equivalent BARD1 and BRCA1-binding proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA  
10       technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test mutants in order to examine DNA binding activity at the molecular level.

15           One may also prepare fusion proteins and peptides, *e.g.*, where the BARD1 or BRCA1-binding protein coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions,  
20       respectively).

          Encompassed by the invention are DNA segments encoding relatively small peptides, such as, for example, peptides of from about 15 to about 50 amino acids in length, and more preferably, of from about 15 to about 30 amino acids in length; and also larger polypeptides up  
25       to and including proteins corresponding to the full-length sequences set forth in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46.

## B. Recombinant Vectors, Host Cells and Expression

Recombinant vectors form important further aspects of the present invention. The term "expression vector or construct" means any type of genetic construct containing a nucleic acid coding for a gene product in which part or all of the nucleic acid encoding sequence is capable of being transcribed. The transcript may be translated into a protein, but it need not be. Thus, in certain embodiments, expression includes both transcription of a gene and translation of a RNA into a gene product. In other embodiments, expression only includes transcription of the nucleic acid, for example, to generate antisense constructs.

Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full length protein or smaller peptide, is positioned under the transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrases "operatively positioned", "under control" or "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the gene.

The promoter may be in the form of the promoter that is naturally associated with a wild-type, polymorphic or mutant BARD1 gene, or BRCA1 binding protein gene, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR technology, in connection with the compositions disclosed herein (PCR technology is disclosed in U.S. Patent 4,683,202 and U.S. Patent 4,682,195, each incorporated herein by reference).

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a wild-type, polymorphic or mutant BARD1 gene, or a BRCA1 binding protein gene in its natural environment. Such promoters may include

promoters normally associated with other genes, and/or promoters isolated from any other bacterial, viral, eukaryotic, or mammalian cell.

5 Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.* (1989), incorporated herein by reference. The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the  
10 introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides.

At least one module in a promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box,  
15 such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

Additional promoter elements regulate the frequency of transcriptional initiation.  
20 Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to  
25 decline. Depending on the promoter, it appears that individual elements can function either co-operatively or independently to activate transcription.

The particular promoter that is employed to control the expression of a nucleic acid is not believed to be critical, so long as it is capable of expressing the nucleic acid in the targeted cell.  
30 Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell. Generally speaking, such a promoter might include either a human or viral promoter.

Preferred promoters include those derived from HSV, including the HNF1 $\alpha$  promoter. Another preferred embodiment is the tetracycline controlled promoter.

In various other embodiments, the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter and the Rous sarcoma virus long terminal repeat can be used to obtain high-level expression of transgenes. The use of other viral or mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of a transgene is contemplated as well, provided that the levels of expression are sufficient for a given purpose. Tables 2 and 3 below list several elements/promoters which may be employed, in the context of the present invention, to regulate the expression of wild-type, polymorphic or mutant BARD1 gene or a BRCA1 binding protein gene. This list is not intended to be exhaustive of all the possible elements involved in the promotion of transgene expression but, merely, to be exemplary thereof.

Enhancers were originally detected as genetic elements that increased transcription from a promoter located at a distant position on the same molecule of DNA. This ability to act over a large distance had little precedent in classic studies of prokaryotic transcriptional regulation. Subsequent work showed that regions of DNA with enhancer activity are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

Additionally any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of a transgene. Use of a T3, T7 or SP6 cytoplasmic expression system is another possible embodiment. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial

polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

**Table 2 - Promoter and Enhancer Elements**

5

Promoter/Enhancer	References
Immunoglobulin Heavy Chain	Banerji <i>et al.</i> , 1983; Gilles <i>et al.</i> , 1983; Grosschedl and Baltimore, 1985; Atchinson and Perry, 1986, 1987; Imler <i>et al.</i> , 1987; Weinberger <i>et al.</i> , 1984; Kiledjian <i>et al.</i> , 1988; Porton <i>et al.</i> ; 1990
Immunoglobulin Light Chain	Queen and Baltimore, 1983; Picard and Schaffner, 1984
T-Cell Receptor	Luria <i>et al.</i> , 1987; Winoto and Baltimore, 1989; Redondo <i>et al.</i> ; 1990
HLA DQ $\alpha$ and DQ $\beta$	Sullivan and Peterlin, 1987
$\beta$ -Interferon	Goodbourn <i>et al.</i> , 1986; Fujita <i>et al.</i> , 1987; Goodbourn and Maniatis, 1988
Interleukin-2	Greene <i>et al.</i> , 1989
Interleukin-2 Receptor	Greene <i>et al.</i> , 1989; Lin <i>et al.</i> , 1990
MHC Class II 5	Koch <i>et al.</i> , 1989
MHC Class II HLA-DRA	Sherman <i>et al.</i> , 1989
$\beta$ -Actin	Kawamoto <i>et al.</i> , 1988; Ng <i>et al.</i> ; 1989
Muscle Creatine Kinase	Jaynes <i>et al.</i> , 1988; Horlick and Benfield, 1989; Johnson <i>et al.</i> , 1989
Prealbumin (Transthyretin)	Costa <i>et al.</i> , 1988
Elastase I	Omitz <i>et al.</i> , 1987
Metallothionein	Karin <i>et al.</i> , 1987; Culotta and Hamer, 1989



Table 2 - Continued

Promoter/Enhancer	References
Collagenase	Pinkert <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987
Albumin Gene	Pinkert <i>et al.</i> , 1987; Tronche <i>et al.</i> , 1989, 1990
$\alpha$ -Fetoprotein	Godbout <i>et al.</i> , 1988; Campere and Tilghman, 1989
t-Globin	Bodine and Ley, 1987; Perez-Stable and Constantini, 1990
$\beta$ -Globin	Trudel and Constantini, 1987
e-fos	Cohen <i>et al.</i> , 1987
c-HA-ras	Triesman, 1986; Deschamps <i>et al.</i> , 1985
Insulin	Edlund <i>et al.</i> , 1985
Neural Cell Adhesion Molecule (NCAM)	Hirsh <i>et al.</i> , 1990
$\alpha_1$ -Antitrypsin	Latimer <i>et al.</i> , 1990
H2B (TH2B) Histone	Hwang <i>et al.</i> , 1990
Mouse or Type I Collagen	Ripe <i>et al.</i> , 1989
Glucose-Regulated Proteins (GRP94 and GRP78)	Chang <i>et al.</i> , 1989
Rat Growth Hormone	Larsen <i>et al.</i> , 1986
Human Serum Amyloid A (SAA)	Edbrooke <i>et al.</i> , 1989
Troponin I (TN I)	Yutzey <i>et al.</i> , 1989
Platelet-Derived Growth Factor	Pech <i>et al.</i> , 1989
Duchenne Muscular Dystrophy	Klamut <i>et al.</i> , 1990

**Table 2 - Continued**

Promoter/Enhancer	References
SV40	Banerji <i>et al.</i> , 1981; Moreau <i>et al.</i> , 1981; Sleight and Lockett, 1985; Firak and Subramanian, 1986; Herr and Clarke, 1986; Imbra and Karin, 1986; Kadesch and Berg, 1986; Wang and Calame, 1986; Ondek <i>et al.</i> , 1987; Kuhl <i>et al.</i> , 1987; Schaffner <i>et al.</i> , 1988
Polyoma	Swartzendruber and Lehman, 1975; Vasseur <i>et al.</i> , 1980; Katinka <i>et al.</i> , 1980, 1981; Tyndell <i>et al.</i> , 1981; Dandolo <i>et al.</i> , 1983; de Villiers <i>et al.</i> , 1984; Hen <i>et al.</i> , 1986; Satake <i>et al.</i> , 1988; Campbell and Villarreal, 1988
Retroviruses	Kriegler and Botchan, 1982, 1983; Levinson <i>et al.</i> , 1982; Kriegler <i>et al.</i> , 1983, 1984a, b, 1988; Bosze <i>et al.</i> , 1986; Miksicek <i>et al.</i> , 1986; Celander and Haseltine, 1987; Thiesen <i>et al.</i> , 1988; Celander <i>et al.</i> , 1988; Chol <i>et al.</i> , 1988; Reisman and Rotter, 1989
Papilloma Virus	Campo <i>et al.</i> , 1983; Lusky <i>et al.</i> , 1983; Spandidos and Wilkie, 1983; Spalholz <i>et al.</i> , 1985; Lusky and Botchan, 1986; Cripe <i>et al.</i> , 1987; Gloss <i>et al.</i> , 1987; Hirochika <i>et al.</i> , 1987; Stephens and Hentschel, 1987; Glue <i>et al.</i> , 1988
Hepatitis B Virus	Bulla and Siddiqui, 1986; Jameel and Siddiqui, 1986; Shaul and Ben-Levy, 1987; Spandau and Lee, 1988; Vannice and Levinson, 1988
Human Immunodeficiency Virus	Muesing <i>et al.</i> , 1987; Hauber and Cullan, 1988; Jakobovits <i>et al.</i> , 1988; Feng and Holland, 1988; Takebe <i>et al.</i> , 1988; Rosen <i>et al.</i> , 1988; Berkhout <i>et al.</i> , 1989; Laspia <i>et al.</i> , 1989; Sharp and Marciniak, 1989; Braddock <i>et al.</i> , 1989

Table 2 - Continued

Promoter/Enhancer	References
Cytomegalovirus	Weber <i>et al.</i> , 1984; Boshart <i>et al.</i> , 1985; Foecking and Hofstetter, 1986
Gibbon Ape Leukemia Virus	Holbrook <i>et al.</i> , 1987; Quinn <i>et al.</i> , 1989

5

Table 3 - Inducible Elements

Element	Inducer	References
MT II	Phorbol Ester (TFA) Heavy metals	Palmiter <i>et al.</i> , 1982; Haslinger and Karin, 1985; Searle <i>et al.</i> , 1985; Stuart <i>et al.</i> , 1985; Imagawa <i>et al.</i> , 1987, Karin <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987b; McNeall <i>et al.</i> , 1989
MMTV (mouse mammary tumor virus)	Glucocorticoids	Huang <i>et al.</i> , 1981; Lee <i>et al.</i> , 1981; Majors and Varmus, 1983; Chandler <i>et al.</i> , 1983; Lee <i>et al.</i> , 1984; Ponta <i>et al.</i> , 1985; Sakai <i>et al.</i> , 1988
$\beta$ -Interferon	poly(rl)x poly(rc)	Tavernier <i>et al.</i> , 1983
Adenovirus 5 <u>E2</u>	Ela	Imperiale and Nevins, 1984
Collagenase	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987a
Stromelysin	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b
SV40	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b

**Table 3 - Continued**

Element	Inducer	References
Murine MX Gene	Interferon, Newcastle Disease Virus	
GRP78 Gene	A23187	Resendez <i>et al.</i> , 1988
$\alpha$ -2-Macroglobulin	IL-6	Kunz <i>et al.</i> , 1989
Vimentin	Serum	Rittling <i>et al.</i> , 1989
MHC Class I Gene H-2kb	Interferon	Blonar <i>et al.</i> , 1989
HSP70	Ela, SV40 Large T Antigen	Taylor <i>et al.</i> , 1989; Taylor and Kingston, 1990a, b
Proliferin	Phorbol Ester-TPA	Mordacq and Linzer, 1989
Tumor Necrosis Factor	FMA	Hensel <i>et al.</i> , 1989
Thyroid Stimulating Hormone a Gene	Thyroid Hormone	Chatterjee <i>et al.</i> , 1989

As indicated, it is contemplated that one may use any regulatory element to express the  
 5 BARD1, B123, BE2, BE14, BE31 and BE445 genes disclosed by the present invention;  
 however, under certain circumstances it may be desirable to use the innate promoter region  
 associated with the gene of interest to control its expression, such as the BARD1 promoter  
 within the 5' flanking region fo the BARD1 genomic clone, as disclosed in SEQ ID NO:122. As  
 noted above, in most cases, genes are regulated at the level of transcription by regulatory  
 10 elements that are located upstream, or 5', to the genes.

In general, to identify regulatory elements for the gene of interest, one would obtain a  
 genomic DNA segment corresponding to the region located between about 10 to 50 nucleotides  
 up to about 2000 nucleotides or more upstream from the transcriptional start site of the gene, *i.e.*  
 15 the nucleotides between positions -10 and -2000. A convenient method used to obtain such a  
 sequence is to utilize restriction enzyme(s) to excise an appropriate DNA fragment. Restriction

enzyme technology is commonly used in the art and will be generally known to the skilled artisan. For example, one may use a combination of enzymes from the extensive range of known restriction enzymes to digest the genomic DNA. Analysis of the digested fragments would determine which enzyme(s) produce the desired DNA fragment. The desired region may  
5 then be excised from the genomic DNA using the enzyme(s). If desired, one may even create a particular restriction site by genetic engineering for subsequent use in ligation strategies.

Alternatively, one may choose to prepare a series of DNA fragments differentiated by size through the use of a deletion assay with linearized DNA. In such an assay, enzymes are  
10 also used to digest the genomic DNA; however, in this case, the enzymes do not recognize specific sites within the DNA but instead digest the DNA from the free end(s). In this case, a series of size differentiated DNA fragments can be achieved by stopping the enzyme reaction after specified time intervals. Of course, one may also choose to use a combination of both restriction enzyme digestion and deletion assay to obtain the desired DNA fragment(s).

15  
Once the desired DNA fragment has been isolated, its potential to regulate a gene and determine the basic regulatory unit may be examined using any one of several conventional techniques. It is recognized that once the core regulatory region is identified, one may choose to employ a longer sequence which comprises the identified regulatory unit. This is because  
20 although the core region is all that is ultimately required, it is believed that particular advantages accrue, in terms of regulation and level of induction achieved where one employs sequences which correspond to the natural control regions over longer regions, *e.g.* from around 25 or so nucleotides to as many as 1000 to 1500 or so nucleotides in length. The preferred length will be in part determined by the type of expression system used and the results desired.

25  
Numerous methods are known in the art for precisely locating regulatory units within larger DNA sequences. Most conveniently, the desired control sequence is isolated within a DNA fragment(s) which is subsequently modified using DNA synthesis techniques to add restriction site linkers to the fragment(s) termini. This modification readily allows the insertion  
30 of the modified DNA fragment into an expression cassette which contains a reporter gene that confers on its recombinant host cell a readily detectable phenotype that is either expressed or inhibited, as may be the case.

Generally reporter genes encode a polypeptide not otherwise produced by the host cell; or a protein or factor produced by the host cell but at much lower levels; or a mutant form of a polypeptide not otherwise produced by the host cell. Preferably the reporter gene encodes an enzyme which produces a colorimetric or fluorometric change in the host cell which is detectable by *in situ* analysis and is a quantitative or semi-quantitative function of transcriptional activation. Exemplary reporter genes encode esterases, phosphatases, proteases and other proteins detected by activity which generates a chromophore or fluorophore as will be known to the skilled artisan. Two well-known examples of such a reporter genes are *E. coli* beta-galactosidase and chloramphenicol-acetyl-transferase (CAT). Alternatively, a reporter gene may render its host cell resistant to a selection agent. For example, the gene *neo* renders cells resistant to the antibiotic neomycin. It is contemplated that virtually any host cell system compatible with the reporter gene cassette may be used to determine the regulatory unit. Thus mammalian or other eukaryotic cells, insect, bacterial or plant cells may be used.

Once a DNA fragment containing the putative regulatory region is inserted into an expression cassette which is in turn inserted into an appropriate host cell system, using any of the techniques commonly known to those of skill in the art, the ability of the fragment to regulate the expression of the reporter gene is assessed. By using a quantitative reporter assay and analyzing a series of DNA fragments of decreasing size, for example produced by convenient restriction endonuclease sites, or through the actions of enzymes such as BAL31, *E. coli* exonuclease III or mung bean nuclease, and which overlap each other a specific number of nucleotides, one may determine both the size and location of the native regulatory unit.

Of course once the core regulatory unit has been determined, one may choose to modify the regulatory unit by mutating certain nucleotides within the core unit. The effects of these modifications may be analyzed using the same reporter assay to determine whether the modifications either enhance or reduce transcription. Thus key nucleotides within the core regulatory sequence can be identified.

It is recognized that regulatory units often contain both elements that either enhance or inhibit transcription. In the case that a regulatory unit is suspected of containing both types of

elements, one may use competitive DNA mobility shift assays to separately identify each element. Those of skill in the art will be familiar the use of DNA mobility shift assays.

It may also be desirable to modify the identified regulatory unit by adding additional sequences to the unit. The added sequences may include additional enhancers, promoters or even other genes. Thus one may, for example, prepare a DNA fragment that contains the native regulatory elements positioned to regulate one or more copies of the native gene and/or another gene or prepare a DNA fragment which contains not one but multiple copies of the promoter region such that transcription levels of the desired gene are relatively increased.

Turning to the expression of the wild-type, polymorphic or mutant BARD1 proteins, or the BRCA1 binding proteins of the present invention, once a suitable clone or clones have been obtained, whether they be cDNA based or genomic, one may proceed to prepare an expression system. The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known to those of skill in recombinant expression. It is believed that virtually any expression system may be employed in the expression of the proteins of the present invention.

Both cDNA and genomic sequences are suitable for eukaryotic expression, as the host cell will generally process the genomic transcripts to yield functional mRNA for translation into protein. Generally speaking, it may be more convenient to employ as the recombinant gene a cDNA version of the gene. It is believed that the use of a cDNA version will provide advantages in that the size of the gene will generally be much smaller and more readily employed to transfect the targeted cell than will a genomic gene, which will typically be up to an order of magnitude larger than the cDNA gene. However, the inventor does not exclude the possibility of employing a genomic version of a particular gene where desired.

In expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. Preferred embodiments include the SV40 polyadenylation signal and the bovine growth hormone polyadenylation signal, convenient and known to function well in various target cells.

Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

5 A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading  
10 frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

15 It is proposed that wild-type, polymorphic or mutant BARD1 genes, or the genes encoding BRCA1 binding proteins may be co-expressed with BRCA1, wherein the proteins may be co-expressed in the same cell or wherein wild-type, polymorphic or mutant BARD1 genes, or the genes encoding BRCA1 binding proteins may be provided to a cell that already has BRCA1. Co-expression may be achieved by co-transfecting the cell with two distinct recombinant  
20 vectors, each bearing a copy of either the respective DNA. Alternatively, a single recombinant vector may be constructed to include the coding regions for both of the proteins, which could then be expressed in cells transfected with the single vector. In either event, the term "co-expression" herein refers to the expression of both the wild-type, polymorphic or mutant BARD1 genes, or the genes encoding BRCA1 binding proteins and the BRCA1 proteins in the  
25 same recombinant cell.

In addition to co-expression with BRCA1, it is proposed that the wild-type, polymorphic or mutant BARD1 genes, or the genes encoding BRCA1 binding proteins may be co-expressed with genes encoding other selected tumor suppressor proteins or peptides. Tumor suppressor  
30 proteins contemplated for use include, but are not limited to, the retinoblastoma, p53, Wilms tumor (WT-1), DCC, neurofibromatosis type 1 (NF-1), von Hippel-Lindau (VHL) disease tumor suppressor, Maspin, Brush-1, BRCA-2 and the multiple tumor suppressor (MTS) or p16 proteins



or peptides. Further particularly contemplated is co-expression with a selected wild-type version of a selected oncogene. Wild-type oncogenes contemplated for use include, but are not limited to, tyrosine kinases, both membrane-associated and cytoplasmic forms, such as members of the Src family, serine/threonine kinases, such as Mos, growth factor and receptors, such as platelet derived growth factor (PDGF), small GTPases (G proteins) including the ras family and Gs-alpha, cyclin-dependent protein kinases (cdk), members of the myc family members including c-myc, N-myc, and L-myc and bcl-2 and family members.

As used herein, the terms "engineered" and "recombinant" cells are intended to refer to a cell into which an exogenous DNA segment or gene, such as a cDNA or gene encoding a BARD1 or BRCA1 binding protein has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced exogenous DNA segment or gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinant cells include those having an introduced cDNA or genomic gene, and also include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene.

To express a recombinant BARD1 or BRCA1 binding protein, whether mutant or wild-type, in accordance with the present invention one would prepare an expression vector that comprises a wild-type, polymorphic or mutant BARD1-, or a BRCA1 binding protein-encoding nucleic acid under the control of one or more promoters. To bring a coding sequence "under the control of" a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame generally between about 1 and about 50 nucleotides "downstream" of (*i.e.*, 3' of) the chosen promoter. The "upstream" promoter stimulates transcription of the DNA and promotes expression of the encoded recombinant protein. This is the meaning of "recombinant expression" in this context.

Many standard techniques are available to construct expression vectors containing the appropriate nucleic acids and transcriptional/translational control sequences in order to achieve protein or peptide expression in a variety of host-expression systems. Cell types available for expression include, but are not limited to, bacteria, such as *E. coli* and *B. subtilis* transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors.

Certain examples of prokaryotic hosts are *E. coli* strain RR1, *E. coli* LE392, *E. coli* B, *E. coli* X 1776 (ATCC No. 31537) as well as *E. coli* W3110 (F-, lambda-, prototrophic, ATCC No. 273325); bacilli such as *Bacillus subtilis*; and other enterobacteriaceae such as *Salmonella typhimurium*, *Serratia marcescens*, and various *Pseudomonas* species.

In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is often transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species. pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of its own proteins.

In addition, phage vectors containing replicon and control sequences that are compatible with the host microorganism can be used as transforming vectors in connection with these hosts. For example, the phage lambda GEM<sup>TM</sup>-11 may be utilized in making a recombinant phage vector which can be used to transform host cells, such as *E. coli* LE392.

Further useful vectors include pIN vectors (Inouye *et al.*, 1985); and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. Other suitable fusion proteins are those with  $\beta$ -galactosidase, ubiquitin, the like.

Promoters that are most commonly used in recombinant DNA construction include the  $\beta$ -lactamase (penicillinase), lactose and tryptophan (*trp*) promoter systems. While these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling those of skill in the art to ligate them functionally with plasmid vectors.

The following details concerning recombinant protein production in bacterial cells, such as *E. coli*, are provided by way of exemplary information on recombinant protein production in general, the adaptation of which to a particular recombinant expression system will be known to those of skill in the art.

5

Bacterial cells, for example, *E. coli*, containing the expression vector are grown in any of a number of suitable media, for example, LB. The expression of the recombinant protein may be induced, *e.g.*, by adding IPTG to the media or by switching incubation to a higher temperature. After culturing the bacteria for a further period, generally of between 2 and 10 24 hours, the cells are collected by centrifugation and washed to remove residual media.

The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the dense inclusion bodies and cell membranes from the soluble cell components. This centrifugation can be performed under conditions whereby the dense 15 inclusion bodies are selectively enriched by incorporation of sugars, such as sucrose, into the buffer and centrifugation at a selective speed.

If the recombinant protein is expressed in the inclusion bodies, as is the case in many instances, these can be washed in any of several solutions to remove some of the contaminating 20 host proteins, then solubilized in solutions containing high concentrations of urea (*e.g.* 8M) or chaotropic agents such as guanidine hydrochloride in the presence of reducing agents, such as  $\beta$ -mercaptoethanol or DTT (dithiothreitol).

Under some circumstances, it may be advantageous to incubate the protein for several 25 hours under conditions suitable for the protein to undergo a refolding process into a conformation which more closely resembles that of the native protein. Such conditions generally include low protein concentrations, less than 500 mg/ml, low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulfide bonds within the 30 protein molecule.

The refolding process can be monitored, for example, by SDS-PAGE, or with antibodies specific for the native molecule (which can be obtained from animals vaccinated with the native molecule or smaller quantities of recombinant protein). Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

For expression in *Saccharomyces*, the plasmid YRp7, for example, is commonly used. This plasmid already contains the *trp1* gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1. The presence of the *trp1* lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination.

Other suitable promoters, which have the additional advantage of transcription controlled by growth conditions, include the promoter region for alcohol dehydrogenase 2, isocytchrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

In addition to micro-organisms, cultures of cells derived from multicellular organisms may also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. In addition to mammalian cells, these include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus); and plant cell

systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing one or more wild-type, polymorphic or mutant BARD1, or BRCA1 binding protein coding sequences.

5

In a useful insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The wild-type, polymorphic or mutant BARD1 coding sequences or the BRCA1 binding protein coding sequences are cloned into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequences results in the inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (e.g., U.S. Patent No. 4,215,051, Smith, incorporated herein by reference).

10

15

Examples of useful mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, COS-7, 293, HepG2, 3T3, RIN and MDCK cell lines. In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein.

20

25

Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cells lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells such as 293 cells have already been shown to produce active BARD1.

30

Expression vectors for use in mammalian such cells ordinarily include an origin of replication (as necessary), a promoter located in front of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences. The origin of replication may be provided either by construction of the

vector to include an exogenous origin, such as may be derived from SV40 or other viral (e.g., Polyoma, Adeno, VSV, BPV) source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient.

5

The promoters may be derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter). Further, it is also possible, and may be desirable, to utilize promoter or control sequences normally associated with the desired wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein gene sequence, provided such control sequences are compatible with the host cell systems.

10

A number of viral based expression systems may be utilized, for example, commonly used promoters are derived from polyoma, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication. Smaller or larger SV40 fragments may also be used, provided there is included the approximately 250 bp sequence extending from the *HindIII* site toward the *BglII* site located in the viral origin of replication.

15

20

In cases where an adenovirus is used as an expression vector, the coding sequences may be ligated to an adenovirus transcription/ translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins in infected hosts.

25

Specific initiation signals may also be required for efficient translation of wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may additionally need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary

30

signals. It is well known that the initiation codon must be in-frame (or in-phase) with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators.

In eukaryotic expression, one will also typically desire to incorporate into the transcriptional unit an appropriate polyadenylation site (*e.g.*, 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

For long-term, high-yield production of recombinant wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins, stable expression is preferred. For example, cell lines that stably express constructs encoding wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with vectors controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines.

A number of selection systems may be used, including, but not limited, to the herpes simplex virus thymidine kinase, hypoxanthine-guanine phosphoribosyltransferase and adenine phosphoribosyltransferase genes, in tk-, hgpri- or apri- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, that confers resistance to methotrexate; gpt, that confers resistance to mycophenolic acid; neo, that confers resistance to the aminoglycoside G-418; and hygri, that confers resistance to hygromycin.

Animal cells can be propagated *in vitro* in two modes: as non-anchorage dependent cells growing in suspension throughout the bulk of the culture or as anchorage-dependent cells requiring attachment to a solid substrate for their propagation (*i.e.*, a monolayer type of cell growth).

5

Non-anchorage dependent or suspension cultures from continuous established cell lines are the most widely used means of large scale production of cells and cell products. However, suspension cultured cells have limitations, such as tumorigenic potential and lower protein production than adherent cells.

10

Large scale suspension culture of mammalian cells in stirred tanks is a common method for production of recombinant proteins. Two suspension culture reactor designs are in wide use - the stirred reactor and the airlift reactor. The stirred design has successfully been used on an 8000 liter capacity for the production of interferon. Cells are grown in a stainless steel tank with a height-to-diameter ratio of 1:1 to 3:1. The culture is usually mixed with one or more agitators, based on bladed disks or marine propeller patterns. Agitator systems offering less shear forces than blades have been described. Agitation may be driven either directly or indirectly by magnetically coupled drives. Indirect drives reduce the risk of microbial contamination through seals on stirrer shafts.

20

The airlift reactor, also initially described for microbial fermentation and later adapted for mammalian culture, relies on a gas stream to both mix and oxygenate the culture. The gas stream enters a riser section of the reactor and drives circulation. Gas disengages at the culture surface, causing denser liquid free of gas bubbles to travel downward in the downcomer section of the reactor. The main advantage of this design is the simplicity and lack of need for mechanical mixing. Typically, the height-to-diameter ratio is 10:1. The airlift reactor scales up relatively easily, has good mass transfer of gases and generates relatively low shear forces.

25

It is contemplated that the wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins of the invention may be "overexpressed", *i.e.*, expressed in increased levels relative to its natural expression in cells. Such overexpression may be assessed by a variety of methods, including radio-labelling and/or protein purification. However, simple and direct methods are

30



preferred, for example, those involving SDS/PAGE and protein staining or western blotting, followed by quantitative analyses, such as densitometric scanning of the resultant gel or blot. A specific increase in the level of the recombinant protein or peptide in comparison to the level in natural cells is indicative of overexpression, as is a relative abundance of the specific protein in relation to the other proteins produced by the host cell and, *e.g.*, visible on a gel.

### C. Nucleic Acid Detection

In addition to their use in directing the expression of the wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins, the nucleic acid sequences disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments.

#### 1. Hybridization

The use of a hybridization probe of between 17 and 100 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 20 bases in length are generally preferred, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of particular hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having stretches of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of genes or RNAs or to provide primers for amplification of DNA or RNA from tissues. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence.

For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high

temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Such high stringency conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating specific genes or detecting specific mRNA transcripts. It is generally appreciated that conditions  
5 can be rendered more stringent by the addition of increasing amounts of formamide.

For certain applications, for example, substitution of nucleotides by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not  
10 perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Thus, hybridization  
15 conditions can be readily manipulated depending on the desired results.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 1.0 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include  
20 approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, at temperatures ranging from approximately 40°C to about 72°C.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining  
25 hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates  
30 are known that can be employed to provide a detection means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization, as in PCR, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface to remove non-specifically bound probe molecules, hybridization is detected, or even quantified, by means of the label.

## 2. Amplification and PCR

Nucleic acid used as a template for amplification is isolated from cells contained in the biological sample, according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA. In one embodiment, the RNA is whole cell RNA and is used directly as the template for amplification.

Pairs of primers that selectively hybridize to nucleic acids corresponding to wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein are contacted with the isolated nucleic acid under conditions that permit selective hybridization. The term "primer", as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of

amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax technology).

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and each incorporated herein by reference in entirety.

Briefly, in PCR, two primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, *e.g.*, *Taq* polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641, filed December 21, 1990, incorporated herein by reference. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction ("LCR"), disclosed in EPA No. 320 308, incorporated herein by reference in its entirety. In LCR, two complementary probe

pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, incorporated herein by reference, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template- and enzyme-dependent synthesis. The primers may be modified by labelling with a capture moiety (e.g., biotin) and/or a detector moiety (e.g., enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference. In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by an RNA polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into single stranded DNA, which is then converted to double stranded DNA, and then transcribed once again with an RNA polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey *et al.*, EPA No. 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an

5 RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

15 Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, M.A., In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y., 1990 incorporated by reference).

20 Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention.

25 Following any amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook *et al.*, 1989.

30 Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption,

partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography.

Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, and the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and can be found in many standard books on molecular protocols. See Sambrook *et al.*, 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.



All the essential materials and reagents required for detecting wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein markers in a biological sample may be assembled together in a kit. This generally will comprise preselected primers for specific markers. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (RT, Taq, *etc.*), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification.

Such kits generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker primer pair. Preferred pairs of primers for amplifying nucleic acids are selected to amplify the sequences specified in SEQ ID NO:1, SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130.

In another embodiment, such kits will comprise hybridization probes specific for wild-type, polymorphic or mutant BARD1 or for BRCA1 binding protein chosen from a group including nucleic acids corresponding to the sequences specified in SEQ ID NO:1, any one of SEQ ID NO:9 through SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130. Such kits generally will comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker hybridization probe.

### 3. Other Assays

Other methods for genetic screening to accurately detect mutations in genomic DNA, cDNA or RNA samples may be employed, depending on the specific situation. When screening for mutations in the genomic DNA, it will be preferable to use probes or primers from intronic

sequences, such as the intronic sequences disclosed herein for the BARD1 gene in SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 and SEQ ID NO:130. In particular, mutations which are weakly expressed or are not expressed at all will still be able to be detected in the germline genomic DNA using intronic probes. Additionally, mutations which effect the splice sites of the gene will be able to be detected using intronic sequences, especially, as is the case with the BARD1 gene disclosed herein, when the intron/exon borders have been defined. This is the case for each of the eleven exons of the BARD1 gene, contained within the genomic contigs disclosed in SEQ ID NO:122 (exon I, bp 2031-2188), SEQ ID NO:123 (exon II, bp 2623-2679; exon III, bp 5421-6415), SEQ ID NO:124 (exon IV, bp 621-1570), SEQ ID NO:125 (exon V, bp 451-5318), SEQ ID NO:126 (exon VI, bp 508-680), SEQ ID NO:127 (exon VII, bp 548-656), SEQ ID NO:128 (exon VIII, bp 566-698), SEQ ID NO:129 (exon IX, bp 226-318), and SEQ ID NO:130 (exon X, bp 519-616; exon XI, bp 2019-2351).

Historically, a number of different methods have been used to detect point mutations, including denaturing gradient gel electrophoresis ("DGGE"), restriction enzyme polymorphism analysis, chemical and enzymatic cleavage methods, and others. The more common procedures currently in use include direct sequencing of target regions amplified by PCR<sup>TM</sup> (see above) and single-strand conformation polymorphism analysis ("SSCP").

Another method of screening for point mutations is based on RNase cleavage of base pair mismatches in RNA/DNA and RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single and multiple base point mutations.

U.S. Patent No. 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. After the RNase cleavage reaction, the RNase is inactivated by proteolytic digestion and organic extraction, and the cleavage products are denatured by heating and analyzed by electrophoresis on denaturing polyacrylamide gels.

For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as +.

5

Currently available RNase mismatch cleavage assays, including those performed according to U.S. Patent No. 4,946,773, require the use of radiolabeled RNA probes. Myers and Maniatis in U.S. Patent No. 4,946,773 describe the detection of base pair mismatches using RNase A. Other investigators have described the use of *E. coli* enzyme, RNase I, in mismatch assays. Because it has broader cleavage specificity than RNase A, RNase I would be a desirable enzyme to employ in the detection of base pair mismatches if components can be found to decrease the extent of non-specific cleavage and increase the frequency of cleavage of mismatches. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is shown in their literature to cleave three out of four known mismatches, provided the enzyme level is sufficiently high.

10

15

The RNase protection assay was first used to detect and map the ends of specific mRNA targets in solution. The assay relies on being able to easily generate high specific activity radiolabeled RNA probes complementary to the mRNA of interest by *in vitro* transcription. Originally, the templates for *in vitro* transcription were recombinant plasmids containing bacteriophage promoters. The probes are mixed with total cellular RNA samples to permit hybridization to their complementary targets, then the mixture is treated with RNase to degrade excess unhybridized probe. Also, as originally intended, the RNase used is specific for single-stranded RNA, so that hybridized double-stranded probe is protected from degradation. After inactivation and removal of the RNase, the protected probe (which is proportional in amount to the amount of target mRNA that was present) is recovered and analyzed on a polyacrylamide gel.

20

25

30

The RNase Protection assay was adapted for detection of single base mutations. In this type of RNase A mismatch cleavage assay, radiolabeled RNA probes transcribed *in vitro* from wild-type sequences, are hybridized to complementary target regions derived from test samples. The test target generally comprises DNA (either genomic DNA or DNA amplified by cloning in

plasmids or by PCR<sup>TM</sup>), although RNA targets (endogenous mRNA) have occasionally been used. If single nucleotide (or greater) sequence differences occur between the hybridized probe and target, the resulting disruption in Watson-Crick hydrogen bonding at that position ("mismatch") can be recognized and cleaved in some cases by single-strand specific ribonuclease. To date, RNase A has been used almost exclusively for cleavage of single-base mismatches, although RNase I has recently been shown as useful also for mismatch cleavage. There are recent descriptions of using the MutS protein and other DNA-repair enzymes for detection of single-base mismatches.

#### D. Mutagenesis

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art. As will be appreciated, the technique typically employs a bacteriophage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector which includes within its sequence a DNA sequence encoding the desired protein. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded DNA preparation, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected gene using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of genes may be obtained. For example, recombinant vectors encoding the desired gene may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

## II. BARD1 and BRCA1 Binding Proteins and Peptides

In addition to its ability to bind BRCA1 *in vivo* and *in vitro*, BARD1 shares sequence homology with the two most conserved regions of BRCA1 - the amino-terminal RING motif and the carboxy-terminal BRCT domains. Although the functional properties of the RING domain have not been clearly defined, this motif is found in a variety of proteins that regulate cell growth, including the products of tumor suppressor genes and dominant proto-oncogenes (Saurin *et al.*, 1996).

Several different subgroups of RING proteins are now recognized. The largest of these, which includes BRCA1, features an isolated RING domain that typically resides near the amino-terminus. In other proteins, however, the RING domain forms one element of a tripartite motif that also contains a distinct zinc-binding domain (the B box) and a potential  $\alpha$ -helical coiled-coiled sequence. The RING domain of BARD1 is not found in association with a B-box or coiled-coiled sequence, and in this respect it resembles the isolated RING motif encoded by

*BRCA1*. On the other hand, *BARD1* may represent a novel subgroup within the RING protein family as it is the only known member which contains ankyrin repeats.

Ankyrin repeats are found in a broad spectrum of functionally diverse proteins, and in some instances they have been implicated as sites of highly specific protein-protein interaction (Murre *et al.*, 1989). Although the ankyrin sequences of *BARD1* may serve a similar function, this invention indicates that they are not required for binding to *BRCA1*. Instead, the sequences of *BARD1* and *BRCA1* that mediate their association appear to reside within or nearby their respective RING motifs.

The present invention shows that the ability to interact with *BRCA1* was retained by a segment of *BARD1* (residues 26-142) that includes its RING motif (residues 46-90) but lacks the ankyrin repeats (residues 427-525). Likewise, the interacting sequences of *BRCA1* were localized to the amino-terminal 101 residues, a segment of the protein that also encompasses the RING motif (residues 20-68).

It has been proposed that one possible function of the RING domain would be to provide a surface for protein-protein interactions (Saurin *et al.*, 1996). In support of this notion, *BARD1* does not interact with *BRCA1* polypeptides that have substitutions of amino acids C61 or C64 (FIG. 5A and FIG. 5B), two of the conserved cysteine residues in the RING domain that presumably participate in zinc coordination. This suggests that *BARD1/BRCA1* association is mediated, at least in part, by the RING domain of *BRCA1*. The results are also consistent with a direct heteromeric interaction between the RING domains of *BRCA1* and *BARD1*, although other examples of RING/RING dimerization have not yet been described (Saurin *et al.*, 1996).

The minimal segment of *BRCA1* that successfully bound *BARD1* was comprised of residues 1-101. However, a smaller *BRCA1* segment (residues 1-71) did not interact with *BARD1* despite the fact that it also includes the intact RING motif (residues 20-68). Thus, *BARD1* binding may require multiple points of contact on *BRCA1*, including sequences within the *BRCA1* RING domain and sequences on its carboxy-terminal flank (*i.e.*, residues 72-101).

In any event, BRCA1/BARD1 association appears to be highly specific. The yeast two-hybrid screens with the RING sequences of BRCA1 and BARD1 have not uncovered additional interacting RING proteins, and direct assays of binding between BRCA1 or BARD1 and select members of the RING family have also failed to show evidence of other RING/RING interactions.

A surprising feature of BARD1 is its homology with sequences that lie near the carboxy-terminus of BRCA1. Comparisons of the mouse and human counterparts of BRCA1 have established that this sequence is especially well conserved from an evolutionary standpoint, and the existence of a homologous sequence within BARD1 suggests that it constitutes a discrete amino acid motif with an important but as yet unknown function.

Recently, Koonin *et al.* (1996) reported that this region of BRCA1 is homologous to sequences that reside near the carboxy-termini of the mammalian 53BP1 and yeast RAD9 proteins. Moreover, they also showed that the conserved sequences includes two tandem copies of a novel protein motif - the BRCA1 carboxy-terminal ("BRCT") domain. The function of this motif is not known. Significantly, however, the majority of tumorigenic *BRCA1* lesions associated with familial breast cancer result in mutation or deletion of one or both BRCT domains. Thus, these motifs are likely to play a crucial role in BRCA1-mediated tumor suppression. In view of the fact that BRCA1 and BARD1 form a stable complex *in vivo*, it is proposed that the tumor suppressor function of BRCA1 is mediated by the combined activities of the BRCT motifs from both proteins.

The present invention therefore provides purified, and in preferred embodiments, substantially purified, BARD1 and BRCA1 binding proteins and peptides. The term "purified BARD1 and BRCA1 binding protein or peptide" as used herein, is intended to refer to a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding proteinaceous composition, isolatable from mammalian cells or recombinant host cells, wherein the wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide is purified to any degree relative to its naturally-obtainable state, *i.e.*, relative to its purity within a cellular extract. A purified wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide therefore also

refers to a wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide free from the environment in which it naturally occurs.

Wild-type, polymorphic or mutant BARD1 proteins may be full length proteins, such as being 777, 770 or 752 amino acids in length. Wild-type, polymorphic or mutant BARD1 proteins, polypeptides and peptides may also be less than full length proteins, such as individual domains, regions or even epitopic peptides. Where less than full length wild-type, polymorphic or mutant BARD1 proteins are concerned the most preferred will be those containing predicted immunogenic sites and those containing the functional domains identified herein.

For example, wild-type, polymorphic or mutant BARD1 protein domains consisting essentially of an amino-terminal RING motif or domain; an ankyrin repeat region or regions; or a carboxy-terminal BRCT domain or domains may be prepared. Preferred wild-type, polymorphic or mutant BARD1 protein domains or fragments will be those sufficient to bind to BRCA1, as exemplified by a BRCA1 binding domain that comprises the sequence of residues 26-142 from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39, and which binds to the BRCA1 protein.

Generally, "purified" will refer to a wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide composition that has been subjected to fractionation to remove various non-wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide components, and which composition substantially retains its wild-type, polymorphic or mutant BARD1 or BRCA1 binding activity, as may be assessed by binding to BRCA1 and forming complexes with BRCA1.

Where the term "substantially purified" is used, this will refer to a composition in which the wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide forms the major component of the composition, such as constituting about 50% of the proteins in the composition or more. In preferred embodiments, a substantially purified protein will constitute more than 60%, 70%, 80%, 90%, 95%, 99% or even more of the proteins in the composition.



A polypeptide or protein that is "purified to homogeneity," as applied to the present invention, means that the polypeptide or protein has a level of purity where the polypeptide or protein is substantially free from other proteins and biological components. For example, a purified polypeptide or protein will often be sufficiently free of other protein components so that degradative sequencing may be performed successfully.

Various methods for quantifying the degree of purification of wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins or peptides will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific BRCA1 binding activity of a fraction, or assessing the number of polypeptides within a fraction by gel electrophoresis. Assessing the number of polypeptides within a fraction by SDS/PAGE analysis will often be preferred in the context of the present invention as this is straightforward.

To purify a wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide a natural or recombinant composition comprising at least some wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins or peptides will be subjected to fractionation to remove various non-wild-type, polymorphic or mutant BARD1 or BRCA1 binding components from the composition. Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulfate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite, lectin affinity and other affinity chromatography steps; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques.

A specific example presented herein is the purification of a BARD1 fusion protein using a specific binding partner. Such purification methods are routine in the art. As the present invention provides DNA sequences for BARD1 proteins, any fusion protein purification method can now be practiced. This is currently exemplified by the generation of a BARD1-glutathione S-transferase fusion protein, expression in *E. coli*, and isolation to homogeneity using affinity chromatography on glutathione-agarose.

The exemplary purification method disclosed herein represents one method to prepare a substantially purified wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide. This method is preferred as it results in the substantial purification of the wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide in yields sufficient for further characterization and use. However, given the DNA and proteins provided by the present invention, any purification method can now be employed.

Although preferred for use in certain embodiments, there is no general requirement that the wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein or peptide always be provided in their most purified state. Indeed, it is contemplated that less substantially purified wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins or peptides, which are nonetheless enriched in wild-type, polymorphic or mutant BARD1 or BRCA1 binding protein compositions, relative to the natural state, will have utility in certain embodiments. These include, for example, binding to BRCA1, as may be used to purify BRCA1; and antibody generation where subsequent screening assays using purified wild-type, polymorphic or mutant BARD1 or BRCA1 binding proteins are conducted.

Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein. Inactive products also have utility in certain embodiments, such as, *e.g.*, in antibody generation.

### III. Antibodies to BARD1 and Other BRCA1 Binding Proteins

#### A. Epitopic Core Sequences

Peptides corresponding to one or more antigenic determinants, or "epitopic core regions", of wild-type, polymorphic or mutant BARD1 and the other BRCA1-binding proteins of the present invention can also be prepared. Such peptides should generally be at least five or six amino acid residues in length, will preferably be about 10, 15, 20, 25 or about 30 amino acid residues in length, and may contain up to about 35-50 residues or so.

Synthetic peptides will generally be about 35 residues long, which is the approximate upper length limit of automated peptide synthesis machines, such as those available from Applied Biosystems (Foster City, CA). Longer peptides may also be prepared, e.g., by recombinant means.

5

U.S. Patent 4,554,101, (Hopp) incorporated herein by reference, teaches the identification and preparation of epitopes from primary amino acid sequences on the basis of hydrophilicity. Through the methods disclosed in Hopp, one of skill in the art would be able to identify epitopes from within an amino acid sequence such as the wild-type, polymorphic or mutant BARD1 sequences disclosed herein (SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39) and the other BRCA1-binding proteins encoded by the isolated nucleic acid sequences of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46.

10  
15

Numerous scientific publications have also been devoted to the prediction of secondary structure, and to the identification of epitopes, from analyses of amino acid sequences (Chou & Fasman, 1974a,b; 1978a,b, 1979). Any of these may be used, if desired, to supplement the teachings of Hopp in U.S. Patent 4,554,101.

20

Moreover, computer programs are currently available to assist with predicting antigenic portions and epitopic core regions of proteins. Examples include those programs based upon the Jameson-Wolf analysis (Jameson & Wolf, 1998; Wolf *et al.*, 1988), the program PepPlot® (Brutlag *et al.*, 1990; Weinberger *et al.*, 1985), and other new programs for protein tertiary structure prediction (Fetrow & Bryant, 1993). Further commercially available software capable of carrying out such analyses is termed MacVector (IBI, New Haven, CT).

25

In further embodiments, major antigenic determinants of a polypeptide may be identified by an empirical approach in which portions of the gene encoding the polypeptide are expressed in a recombinant host, and the resulting proteins tested for their ability to elicit an immune response. For example, PCR can be used to prepare a range of peptides lacking successively

30

longer fragments of the C-terminus of the protein. The immunoactivity of each of these peptides is determined to identify those fragments or domains of the polypeptide that are immunodominant. Further studies in which only a small number of amino acids are removed at each iteration then allows the location of the antigenic determinants of the polypeptide to be more precisely determined.

Another method for determining the major antigenic determinants of a polypeptide is the SPOTs™ system (Genosys Biotechnologies, Inc., The Woodlands, TX). In this method, overlapping peptides are synthesized on a cellulose membrane, which following synthesis and deprotection, is screened using a polyclonal or monoclonal antibody. The antigenic determinants of the peptides which are initially identified can be further localized by performing subsequent syntheses of smaller peptides with larger overlaps, and by eventually replacing individual amino acids at each position along the immunoreactive peptide.

Once one or more such analyses are completed, polypeptides are prepared that contain at least the essential features of one or more antigenic determinants. The peptides are then employed in the generation of antisera against the polypeptide. Minigenes or gene fusions encoding these determinants can also be constructed and inserted into expression vectors by standard methods, for example, using PCR cloning methodology.

The use of such small peptides for vaccination typically requires conjugation of the peptide to an immunogenic carrier protein, such as hepatitis B surface antigen, keyhole limpet hemocyanin or bovine serum albumin. Methods for performing this conjugation are well known in the art.

## **B. Antibody Generation**

In certain embodiments, the present invention provides antibodies that bind with high specificity to wild-type, polymorphic or mutant BARD1, and other BRCA1 binding proteins provided herein. Thus, antibodies that bind to the protein products of the isolated nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID

NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46 are provided. Antibodies specific for the wild-type and polymorphic proteins and peptides and those specific for any one of a number of mutants are provided. As detailed above, in addition to antibodies generated against the full length proteins, antibodies may also be generated in response to smaller constructs comprising epitopic core regions, including wild-type, polymorphic and mutant epitopes.

As used herein, the term "antibody" is intended to refer broadly to any immunologic binding agent such as IgG, IgM, IgA, IgD and IgE. Generally, IgG and/or IgM are preferred because they are the most common antibodies in the physiological situation and because they are most easily made in a laboratory setting.

Monoclonal antibodies (MAbs) are recognized to have certain advantages, *e.g.*, reproducibility and large-scale production, and their use is generally preferred. The invention thus provides monoclonal antibodies of the human, murine, monkey, rat, hamster, rabbit and even chicken origin. Due to the ease of preparation and ready availability of reagents, murine monoclonal antibodies will often be preferred.

However, "humanized" antibodies are also contemplated, as are chimeric antibodies from mouse, rat, or other species, bearing human constant and/or variable region domains, bispecific antibodies, recombinant and engineered antibodies and fragments thereof. Methods for the development of antibodies that are "custom-tailored" to the patient's tumor are likewise known and such custom-tailored antibodies are also contemplated.

The term "antibody" is used to refer to any antibody-like molecule that has an antigen binding region, and includes antibody fragments such as Fab', Fab, F(ab')<sub>2</sub>, single domain antibodies (DABs), Fv, scFv (single chain Fv), and the like. The techniques for preparing and using various antibody-based constructs and fragments are well known in the art.

Means for preparing and characterizing antibodies are well known in the art (See, *e.g.*, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference).

5           The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic wild-type, polymorphic or mutant BARD1 or other BRCA1 binding protein composition in accordance with the present invention and collecting antisera from that immunized animal.

10

A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

15

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as  
20   ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobenzoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

25

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Suitable adjuvants include all acceptable immunostimulatory compounds, such as cytokines, toxins or synthetic compositions.

30

Adjuvants that may be used include IL-1, IL-2, IL-4, IL-7, IL-12, g-interferon, GMCSF, BCG, aluminum hydroxide, MDP compounds, such as thur-MDP and nor-MDP, CGP (MTP-PE), lipid A, and monophosphoryl lipid A (MPL). RIBI, which contains three components

extracted from bacteria, MPL, trehalose dimycolate (TDM) and cell wall skeleton (CWS) in a 2% squalene/Tween 80 emulsion. MHC antigens may even be used.

Exemplary, often preferred adjuvants include complete Freund's adjuvant (a non-specific  
5 stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

In addition to adjuvants, it may be desirable to coadminister biologic response modifiers (BRM), which have been shown to upregulate T cell immunity or downregulate suppressor cell  
10 activity. Such BRMs include, but are not limited to, Cimetidine (CIM; 1200 mg/d) (Smith/Kline, PA); or low-dose Cyclophosphamide (CYP; 300 mg/m<sup>2</sup>) (Johnson/ Mead, NJ) and Cytokines such as  $\gamma$ -interferon, IL-2, or IL-12 or genes encoding proteins involved in immune helper functions, such as B-7.

15 The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following  
20 immunization.

A second, booster injection, may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be  
25 used to generate MAbs.

For production of rabbit polyclonal antibodies, the animal can be bled through an ear vein or alternatively by cardiac puncture. The removed blood is allowed to coagulate and then centrifuged to separate serum components from whole cells and blood clots. The serum may be  
30 used as is for various applications or else the desired antibody fraction may be purified by well-known methods, such as affinity chromatography using another antibody, a peptide bound to a solid matrix, or by using, e.g., protein A or protein G chromatography.

MAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*,  
5 a purified or partially purified wild-type, polymorphic or mutant BARD1, and other BRCA1 binding protein, polypeptide, peptide or domain, be it a wild-type or mutant composition. The immunizing composition is administered in a manner effective to stimulate antibody producing cells.

10 The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher  
15 percentage of stable fusions.

The animals are injected with antigen, generally as described above. The antigen may be coupled to carrier molecules such as keyhole limpet hemocyanin if necessary. The antigen would typically be mixed with adjuvant, such as Freund's complete or incomplete adjuvant.  
20 Booster injections with the same antigen would occur at approximately two-week intervals.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood  
25 sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible.

Often, a panel of animals will have been immunized and the spleen of animal with the  
30 highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately  $5 \times 10^7$  to  $2 \times 10^8$  lymphocytes.



The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Gefter *et al.* (1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about  $1 \times 10^{-6}$  to  $1 \times 10^{-8}$ . However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is

generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways.

A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion (*e.g.*, a syngeneic mouse). Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide MAbs in high concentration.

The individual cell lines could also be cultured *in vitro*, where the MAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations.

5

MAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography. Fragments of the monoclonal antibodies of the invention can be obtained from the monoclonal antibodies so produced by methods which include digestion with enzymes, such as pepsin or papain, and/or by cleavage of disulfide bonds by chemical reduction. Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer.

10

It is also contemplated that a molecular cloning approach may be used to generate monoclonals. For this, combinatorial immunoglobulin phagemid libraries are prepared from RNA isolated from the spleen of the immunized animal, and phagemids expressing appropriate antibodies are selected by panning using cells expressing the antigen and control cells. The advantages of this approach over conventional hybridoma techniques are that approximately  $10^4$  times as many antibodies can be produced and screened in a single round, and that new specificities are generated by H and L chain combination which further increases the chance of finding appropriate antibodies.

20

Alternatively, monoclonal antibody fragments encompassed by the present invention can be synthesized using an automated peptide synthesizer, or by expression of full-length gene or of gene fragments in *E. coli*.

25

### C. Antibody Conjugates

The present invention further provides antibodies against wild-type, polymorphic or mutant BARD1, and other BRCA1 binding proteins, generally of the monoclonal type, that are linked to one or more other agents to form an antibody conjugate. Any antibody of sufficient selectivity, specificity and affinity may be employed as the basis for an antibody conjugate.

30

Such properties may be evaluated using conventional immunological screening methodology known to those of skill in the art.

5 Certain examples of antibody conjugates are those conjugates in which the antibody is linked to a detectable label. "Detectable labels" are compounds or elements that can be detected due to their specific functional properties, or chemical characteristics, the use of which allows the antibody to which they are attached to be detected, and further quantified if desired. Another such example is the formation of a conjugate comprising an antibody linked to a cytotoxic or anti-cellular agent, as may be termed "immunotoxins". In the context of the present invention, 10 immunotoxins are generally less preferred.

Antibody conjugates are thus preferred for use as diagnostic agents. Antibody diagnostics generally fall within two classes, those for use in *in vitro* diagnostics, such as in a variety of immunoassays, and those for use *in vivo* diagnostic protocols, generally known as 15 "antibody-directed imaging". Again, antibody-directed imaging is less preferred for use with this invention.

Many appropriate imaging agents are known in the art, as are methods for their attachment to antibodies (see, e.g., U.S. patents 5,021,236 and 4,472,509, both incorporated 20 herein by reference). Certain attachment methods involve the use of a metal chelate complex employing, for example, an organic chelating agent such a DTPA attached to the antibody (U.S. Patent 4,472,509). Monoclonal antibodies may also be reacted with an enzyme in the presence of a coupling agent such as glutaraldehyde or periodate. Conjugates with fluorescein markers are prepared in the presence of these coupling agents or by reaction with an isothiocyanate. 25

In the case of paramagnetic ions, one might mention by way of example ions such as chromium (III), manganese (II), iron (III), iron (II), cobalt (II), nickel (II), copper (II), neodymium (III), samarium (III), ytterbium (III), gadolinium (III), vanadium (II), terbium (III), dysprosium (III), holmium (III) and erbium (III), with gadolinium being particularly preferred. 30

Ions useful in other contexts, such as X-ray imaging, include but are not limited to lanthanum (III), gold (III), lead (II), and especially bismuth (III).

In the case of radioactive isotopes for therapeutic and/or diagnostic application, one might mention astatine<sup>211</sup>, <sup>14</sup>carbon, <sup>51</sup>chromium, <sup>36</sup>chlorine, <sup>57</sup>cobalt, <sup>58</sup>cobalt, copper<sup>67</sup>, <sup>152</sup>Eu, gallium<sup>67</sup>, <sup>3</sup>hydrogen, iodine<sup>123</sup>, iodine<sup>125</sup>, iodine<sup>131</sup>, indium<sup>111</sup>, <sup>59</sup>iron, <sup>32</sup>phosphorus, rhenium<sup>186</sup>,  
5 rhenium<sup>188</sup>, <sup>75</sup>selenium, <sup>35</sup>sulphur, technetium<sup>99m</sup> and yttrium<sup>90</sup>. <sup>125</sup>I is often being preferred for use in certain embodiments, and technetium<sup>99m</sup> and indium<sup>111</sup> are also often preferred due to their low energy and suitability for long range detection.

Radioactively labeled monoclonal antibodies of the present invention may be produced  
10 according to well-known methods in the art. For instance, monoclonal antibodies can be iodinated by contact with sodium or potassium iodide and a chemical oxidizing agent such as sodium hypochlorite, or an enzymatic oxidizing agent, such as lactoperoxidase. Monoclonal antibodies according to the invention may be labeled with technetium-<sup>99m</sup> by ligand exchange process, for example, by reducing pertechnetate with stannous solution, chelating the reduced  
15 technetium onto a Sephadex column and applying the antibody to this column or by direct labeling techniques, *e.g.*, by incubating pertechnetate, a reducing agent such as  $\text{SnCl}_2$ , a buffer solution such as sodium-potassium phthalate solution, and the antibody.

Intermediary functional groups which are often used to bind radioisotopes which exist as  
20 metallic ions to antibody are diethylenetriaminepentaacetic acid (DTPA) and ethylene diaminetetracetic acid (EDTA).

Fluorescent labels include rhodamine, fluorescein isothiocyanate and renographin.

25 The much preferred antibody conjugates of the present invention are those intended primarily for use *in vitro*, where the antibody is linked to a secondary binding ligand or to an enzyme (an enzyme tag) that will generate a colored product upon contact with a chromogenic substrate. Examples of suitable enzymes include urease, alkaline phosphatase, (horseradish) hydrogen peroxidase and glucose oxidase. Preferred secondary binding ligands are biotin and  
30 avidin or streptavidin compounds. The use of such labels is well known to those of skill in the art in light and is described, for example, in U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241; each incorporated herein by reference.

#### D. Immunodetection Methods

In still further embodiments, the present invention concerns immunodetection methods for binding, purifying, removing, quantifying or otherwise generally detecting biological components such as wild-type, polymorphic or mutant BARD1, and other BRCA1 binding protein components. The wild-type, polymorphic or mutant BARD1, or other BRCA1 binding proteins or peptides of the present invention may be employed to detect and purify BRCA1, and antibodies prepared in accordance with the present invention, may be employed to detect wild-type, polymorphic or mutant BARD1, or other BRCA1 binding proteins or peptides. As described throughout the present application, the use of wild-type, polymorphic and mutant specific antibodies is contemplated. The steps of various useful immunodetection methods have been described in the scientific literature, such as, *e.g.*, Nakamura *et al.* (1987), incorporated herein by reference.

In general, the immunobinding methods include obtaining a sample suspected of containing a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or peptide, and contacting the sample with a first anti-wild-type, polymorphic or mutant BARD1, or BRCA1 binding protein antibody in accordance with the present invention, as the case may be, under conditions effective to allow the formation of immunocomplexes.

These methods include methods for purifying wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein, as may be employed in purifying wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein from patients' samples or for purifying recombinantly expressed wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein. In these instances, the antibody removes the antigenic wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein component from a sample. The antibody will preferably be linked to a solid support, such as in the form of a column matrix, and the sample suspected of containing the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigenic component will be applied to the immobilized antibody. The unwanted components will be washed from the column, leaving the antigen immunocomplexed to the immobilized antibody, which wild-type, polymorphic or mutant BARD1, or other BRCA1

binding protein antigen is then collected by removing the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein from the column.

5 The immunobinding methods also include methods for detecting or quantifying the amount of a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein reactive component in a sample, which methods require the detection or quantification of any immune complexes formed during the binding process. Here, one would obtain a sample suspected of containing a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or peptide, and contact the sample with an antibody against wild-type, polymorphic or  
10 mutant BARD1, or other BRCA1 binding protein, and then detect or quantify the amount of immune complexes formed under the specific conditions.

In terms of antigen detection, the biological sample analyzed may be any sample that is suspected of containing a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding  
15 protein-specific antigen, such as a breast, ovarian or uterine cancer tissue section or specimen, a homogenized breast, ovarian or uterine cancer tissue extract, a breast, ovarian or uterine cancer cell, separated or purified forms of any of the above wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein-containing compositions, or even any biological fluid that comes into contact with breast, ovarian or uterine cancer tissue, including blood and serum,  
20 although tissue samples and extracts are preferred.

Contacting the chosen biological sample with the antibody under conditions effective and for a period of time sufficient to allow the formation of immune complexes (primary immune complexes) is generally a matter of simply adding the antibody composition to the  
25 sample and incubating the mixture for a period of time long enough for the antibodies to form immune complexes with, *i.e.*, to bind to, any wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigens present. After this time, the sample-antibody composition, such as a tissue section, ELISA plate, dot blot or western blot, will generally be washed to remove any non-specifically bound antibody species, allowing only those antibodies  
30 specifically bound within the primary immune complexes to be detected.

In general, the detection of immunocomplex formation is well known in the art and may be achieved through the application of numerous approaches. These methods are generally based upon the detection of a label or marker, such as any of those radioactive, fluorescent, biological or enzymatic tags. U.S. Patents concerning the use of such labels include 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each incorporated herein by reference. Of course, one may find additional advantages through the use of a secondary binding ligand such as a second antibody or a biotin/avidin ligand binding arrangement, as is known in the art.

The wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibody employed in the detection may itself be linked to a detectable label, wherein one would then simply detect this label, thereby allowing the amount of the primary immune complexes in the composition to be determined.

Alternatively, the first antibody that becomes bound within the primary immune complexes may be detected by means of a second binding ligand that has binding affinity for the antibody. In these cases, the second binding ligand may be linked to a detectable label. The second binding ligand is itself often an antibody, which may thus be termed a "secondary" antibody. The primary immune complexes are contacted with the labeled, secondary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of secondary immune complexes. The secondary immune complexes are then generally washed to remove any non-specifically bound labeled secondary antibodies or ligands, and the remaining label in the secondary immune complexes is then detected.

Further methods include the detection of primary immune complexes by a two step approach. A second binding ligand, such as an antibody, that has binding affinity for the antibody is used to form secondary immune complexes, as described above. After washing, the secondary immune complexes are contacted with a third binding ligand or antibody that has binding affinity for the second antibody, again under conditions effective and for a period of time sufficient to allow the formation of immune complexes (tertiary immune complexes). The third ligand or antibody is linked to a detectable label, allowing detection of the tertiary immune complexes thus formed. This system may provide for signal amplification if this is desired.



The immunodetection methods of the present invention have evident utility in the diagnosis or prognosis of conditions such as breast, ovarian, uterine and other forms of cancer. Here, a biological or clinical sample suspected of containing a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein, peptide or mutant is used. However, these embodiments also have applications to non-clinical samples, such as in the titering of antigen or antibody samples, in the selection of hybridomas, and the like.

In the clinical diagnosis or monitoring of patients with breast, ovarian, uterine and other forms of cancer, the detection of a BARD1 or BRCA1 binding protein mutant, or an alteration in the levels of BARD1 or BRCA1 binding protein, in comparison to the levels in a corresponding biological sample from a normal subject is indicative of a patient with breast, ovarian, uterine or another form of cancer.

However, as is known to those of skill in the art, such a clinical diagnosis would not necessarily be made on the basis of this method in isolation. Those of skill in the art are very familiar with differentiating between significant differences in types or amounts of biomarkers, which represent a positive identification, and low level or background changes of biomarkers. Indeed, background expression levels are often used to form a "cut-off" above which increased detection will be scored as significant or positive.

### 1. ELISAs

As detailed above, immunoassays, in their most simple and direct sense, are binding assays. Certain preferred immunoassays are the various types of enzyme linked immunosorbent assays (ELISAs) and radioimmunoassays (RIA) known in the art. Immunohistochemical detection using tissue sections is also particularly useful. However, it will be readily appreciated that detection is not limited to such techniques, and Western blotting, dot blotting, FACS analyses, and the like may also be used.

In one exemplary ELISA, the anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibodies of the invention are immobilized onto a selected surface

exhibiting protein affinity, such as a well in a polystyrene microtiter plate. Then, a test composition suspected of containing the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigen, such as a clinical sample, is added to the wells. After binding and washing to remove non-specifically bound immune complexes, the bound wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigen may be detected. Detection is generally achieved by the addition of another anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibody that is linked to a detectable label. This type of ELISA is a simple "sandwich ELISA". Detection may also be achieved by the addition of a second anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibody, followed by the addition of a third antibody that has binding affinity for the second antibody, with the third antibody being linked to a detectable label.

In another exemplary ELISA, the samples suspected of containing the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigen are immobilized onto the well surface and then contacted with the anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibodies of the invention. After binding and washing to remove non-specifically bound immune complexes, the bound anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibodies are detected. Where the initial anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibodies are linked to a detectable label, the immune complexes may be detected directly. Again, the immune complexes may be detected using a second antibody that has binding affinity for the first anti-wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibody, with the second antibody being linked to a detectable label.

Another ELISA in which the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding proteins or peptides are immobilized, involves the use of antibody competition in the detection. In this ELISA, labeled antibodies against wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein are added to the wells, allowed to bind, and detected by means of their label. The amount of wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antigen in an unknown sample is then determined by mixing the sample with the labeled antibodies against wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein before or during incubation with coated wells. The presence of wild-type,

polymorphic or mutant BARD1, or other BRCA1 binding protein in the sample acts to reduce the amount of antibody against wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein available for binding to the well and thus reduces the ultimate signal. This is also appropriate for detecting antibodies against wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein in an unknown sample, where the unlabeled antibodies bind to the antigen-coated wells and also reduces the amount of antigen available to bind the labeled antibodies.

Irrespective of the format employed, ELISAs have certain features in common, such as coating, incubating or binding, washing to remove non-specifically bound species, and detecting the bound immune complexes. These are described as follows:

In coating a plate with either antigen or antibody, one will generally incubate the wells of the plate with a solution of the antigen or antibody, either overnight or for a specified period of hours. The wells of the plate will then be washed to remove incompletely adsorbed material. Any remaining available surfaces of the wells are then "coated" with a nonspecific protein that is antigenically neutral with regard to the test antisera. These include bovine serum albumin (BSA), casein and solutions of milk powder. The coating allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

In ELISAs, it is probably more customary to use a secondary or tertiary detection means rather than a direct procedure. Thus, after binding of a protein or antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the biological sample to be tested under conditions effective to allow immune complex (antigen/antibody) formation. Detection of the immune complex then requires a labeled secondary binding ligand or antibody, or a secondary binding ligand or antibody in conjunction with a labeled tertiary antibody or third binding ligand.

"Under conditions effective to allow immune complex (antigen/antibody) formation" means that the conditions preferably include diluting the antigens and antibodies with solutions

such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween. These added agents also tend to assist in the reduction of nonspecific background.

5 The "suitable" conditions also mean that the incubation is at a temperature and for a period of time sufficient to allow effective binding. Incubation steps are typically from about 1 to 2 to 4 hours, at temperatures preferably on the order of 25°C to 27°C, or may be overnight at about 4°C or so.

10 Following all incubation steps in an ELISA, the contacted surface is washed so as to remove non-complexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween, or borate buffer. Following the formation of specific immune complexes between the test sample and the originally bound material, and subsequent washing, the occurrence of even minute amounts of immune complexes may be determined.

15 To provide a detecting means, the second or third antibody will have an associated label to allow detection. Preferably, this will be an enzyme that will generate color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the first or second immune complex with a urease, glucose oxidase, alkaline phosphatase or hydrogen peroxidase-conjugated antibody for a period of time and under  
20 conditions that favor the development of further immune complex formation (e.g., incubation for 2 hours at room temperature in a PBS-containing solution such as PBS-Tween).

25 After incubation with the labeled antibody, and subsequent to washing to remove unbound material, the amount of label is quantified, e.g., by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulfonic acid [ABTS] and H<sub>2</sub>O<sub>2</sub>, in the case of peroxidase as the enzyme label. Quantification is then achieved by measuring the degree of color generation, e.g., using a visible spectra spectrophotometer.

## 2. Immunohistochemistry

The antibodies of the present invention may also be used in conjunction with both fresh-frozen and formalin-fixed, paraffin-embedded tissue blocks prepared for study by immunohistochemistry (IHC). For example, each tissue block consists of 50 mg of residual "pulverized" diabetic tissue. The method of preparing tissue blocks from these particulate specimens has been successfully used in previous IHC studies of various prognostic factors, and is well known to those of skill in the art (Brown *et al.*, 1990; Abbondanzo *et al.*, 1990; Allred *et al.*, 1990).

10

Briefly, frozen-sections may be prepared by rehydrating 50 ng of frozen "pulverized" diabetic tissue at room temperature in phosphate buffered saline (PBS) in small plastic capsules; pelleting the particles by centrifugation; resuspending them in a viscous embedding medium (OCT); inverting the capsule and pelleting again by centrifugation; snap-freezing in -70°C isopentane; cutting the plastic capsule and removing the frozen cylinder of tissue; securing the tissue cylinder on a cryostat microtome chuck; and cutting 25-50 serial sections.

15

Permanent-sections may be prepared by a similar method involving rehydration of the 50 mg sample in a plastic microfuge tube; pelleting; resuspending in 10% formalin for 4 hours fixation; washing/pelleting; resuspending in warm 2.5% agar; pelleting; cooling in ice water to harden the agar; removing the tissue/agar block from the tube; infiltrating and embedding the block in paraffin; and cutting up to 50 serial permanent sections.

20

### E. Immunodetection Kits

25

In still further embodiments, the present invention concerns immunodetection kits for use with the immunodetection methods described above. As the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein antibodies are generally used to detect wild-type, polymorphic or mutant BARD1, or other BRCA1 binding proteins or peptides, the antibodies will preferably be included in the kit. However, kits including both such components may be provided. The immunodetection kits will thus comprise, in suitable container means, a first antibody that binds to a wild-type, polymorphic or mutant BARD1, or other BRCA1

30

binding protein or peptide, and optionally, an immunodetection reagent and further optionally, a wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or peptide.

5 In preferred embodiments, monoclonal antibodies will be used. In certain embodiments, the first antibody that binds to the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or peptide may be pre-bound to a solid support, such as a column matrix or well of a microtitre plate.

10 The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with or linked to the given antibody. Detectable labels that are associated with or attached to a secondary binding ligand are also contemplated. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody.

15 Further suitable immunodetection reagents for use in the present kits include the two-component reagent that comprises a secondary antibody that has binding affinity for the first antibody, along with a third antibody that has binding affinity for the second antibody, the third antibody being linked to a detectable label. As noted above, a number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention.

20 The kits may further comprise a suitably aliquoted composition of the wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or polypeptide, whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay.

25 The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit. The components of the kits may be packaged either in aqueous media or in lyophilized form.

30 The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the antibody may be placed, and preferably, suitably aliquoted. Where wild-type, polymorphic or mutant BARD1, or other BRCA1 binding protein or a second or third binding ligand or additional component is provided, the kit will also

generally contain a second, third or other additional container into which this ligand or component may be placed. The kits of the present invention will also typically include a means for containing the antibody, antigen, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into  
5 which the desired vials are retained.

#### IV. Biological Functional Equivalents

As modifications and changes may be made in the structure of wild-type, polymorphic or  
10 mutant BARD1 or the other BRCA1-binding genes and proteins of the present invention, and still obtain molecules having like or otherwise desirable characteristics, such biologically functional equivalents are also encompassed within the present invention.

For example, certain amino acids may be substituted for other amino acids in a protein  
15 structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies, binding sites on substrate molecules or receptors, DNA binding sites, BRCA1-binding regions, or such like. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence (or, of course, its  
20 underlying DNA coding sequence) and nevertheless obtain a protein with like (agonistic) properties. It is thus contemplated by the inventors that various changes may be made in the sequence of wild-type, polymorphic or mutant BARD1 or other BRCA1-binding proteins or peptides, or underlying DNA, without appreciable loss of their biological utility or activity.

25 Equally, the same considerations may be employed to create a protein or peptide with countervailing, e.g., antagonistic properties. This is relevant to the present invention in which BARD1 or other BRCA1-binding mutants or analogues may be generated. For example, a BARD1 or other BRCA1-binding mutant may be generated and tested for BRCA1 binding activity to identify those residues important for BRCA1 and/or DNA binding. BARD1 or other  
30 BRCA1-binding mutants may also be synthesized to reflect a BARD1 or other BRCA1-binding mutant that occurs in the human population and that is linked to the development of breast, ovarian or uterine cancer. Such mutant proteins are particularly contemplated for use in

generating mutant-specific antibodies and such mutant DNA segments may be used as mutant-specific probes and primers.

In terms of functional equivalents, it is well understood by the skilled artisan that, inherent in the definition of a "biologically functional equivalent protein or peptide or gene", is the concept that there is a limit to the number of changes that may be made within a defined portion of the molecule and still result in a molecule with an acceptable level of equivalent biological activity. Biologically functional equivalent peptides are thus defined herein as those peptides in which certain, not most or all, of the amino acids may be substituted.

In particular, where shorter length peptides, such as RING motifs are concerned, it is contemplated that fewer amino acids should be made within the given peptide. Longer domains may have an intermediate number of changes. The full length protein will have the most tolerance for a larger number of changes. Of course, a plurality of distinct proteins/peptides with different substitutions may easily be made and used in accordance with the invention.

It is also well understood that where certain residues are shown to be particularly important to the biological or structural properties of a protein or peptide, *e.g.*, residues in binding regions or active sites, such residues may not generally be exchanged. This is an important consideration in the present invention, where changes in the BRCA1-binding region, the RING motif and the BRCT domains should be carefully considered and subsequently tested to ensure maintenance of biological function, where maintenance of biological function is desired. In this manner, functional equivalents are defined herein as those peptides which maintain a substantial amount of their native biological activity.

Amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. An analysis of the size, shape and type of the amino acid side-chain substituents reveals that arginine, lysine and histidine are all positively charged residues; that alanine, glycine and serine are all a similar size; and that phenylalanine, tryptophan and tyrosine all have a generally similar shape. Therefore, based upon these considerations, arginine, lysine and histidine;



alanine, glycine and serine; and phenylalanine, tryptophan and tyrosine; are defined herein as biologically functional equivalents.

To effect more quantitative changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte & Doolittle, 1982, incorporated herein by reference). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those which are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biological functional equivalent protein or peptide thereby created is intended for use in immunological embodiments, as in certain embodiments of the present invention. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.* with a biological property of the protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm$  1); glutamate (+3.0  $\pm$  1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5  $\pm$  1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those which are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

5

While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes may be effected by alteration of the encoding DNA; taking into consideration also that the genetic code is degenerate and that two or more codons may code for the same amino acid. A table of amino acids and their codons is presented herein for use in such embodiments, as well as for other uses, such as in the design of probes and primers and the like.

10

In addition to the wild-type, polymorphic or mutant BARD1 or other BRCA1 binding peptidyl compounds described herein, the inventors also contemplate that other sterically similar compounds may be formulated to mimic the key portions of the peptide structure or to interact specifically with BRCA1. Such compounds, which may be termed peptidomimetics, may be used in the same manner as the peptides of the invention and hence are also functional equivalents.

15

Certain mimetics that mimic elements of protein secondary structure are described in Johnson *et al.* (1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orientate amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is thus designed to permit molecular interactions similar to the natural molecule.

20

25

Some successful applications of the peptide mimetic concept have focused on mimetics of  $\beta$ -turns within proteins, which are known to be highly antigenic. Likely  $\beta$ -turn structure within a polypeptide can be predicted by computer-based algorithms, as discussed herein. Once the component amino acids of the turn are determined, mimetics can be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

30

The generation of further structural equivalents or mimetics may be achieved by the techniques of modeling and chemical design known to those of skill in the art. The art of receptor modeling is now well known, and by such methods a chemical that binds to wild-type, polymorphic or mutant BARD1 or other BRCA1-binding protein or to a BRCA1-wild-type, polymorphic or mutant BARD1 or other BRCA1-binding protein complex can be designed and then synthesized. It will be understood that all such sterically designed constructs fall within the scope of the present invention.

## V. BRCA1 Binding, Purification and Assays

Certain aspects of this invention concern methods for conveniently evaluating candidate substances to identify compounds capable of stimulating BRCA1 binding to wild-type, polymorphic or mutant BARD1 or other BRCA1 binding protein, or even transcription of wild-type, polymorphic or mutant BARD1 or other BRCA1 binding protein.

Successful candidate substances may function in the absence of mutations in BARD1 or another BRCA1 binding protein, in which case the candidate compound may be termed a "positive stimulator" of BARD1 or the other BRCA1 binding protein. Alternatively, such compounds may stimulate transcription in the presence of mutated BARD1 or another BRCA1 binding protein, overcoming the effects of the mutation, *i.e.*, function to oppose BARD1- or other BRCA1 binding protein-mutant mediated cancer, and thus may be termed "a BARD1 or other BRCA1 binding protein mutant agonist". Compounds may even be discovered which combine both of these actions. Compounds of any such class will likely be useful therapeutic agents for use in treating cancer.

As BARD1 and the other BRCA1 binding proteins are herein shown to bind BRCA1, one method by which to identify a candidate substance capable of stimulating BARD1 or other BRCA1 binding protein is based upon specific protein:protein binding. Accordingly, to conduct such an assay, one may prepare a protein with a BRCA1 binding domain and determine the ability of a candidate substance to increase binding to BRCA1.

As BARD1 and the other BRCA1 binding proteins are also believed to bind DNA, most likely in the context of a complex with BRCA1, another method by which to identify a candidate substance capable of stimulating BARD1 and the other BRCA1 binding proteins is based upon specific protein:DNA binding. Accordingly, to conduct such an assay, one would  
5 prepare a BARD1 or other BRCA1 binding protein and a BRCA1 protein and determine the ability of a candidate substance to increase their binding to a specific DNA segment, *i.e.*, to increase the amount or the binding affinity of a specific protein:DNA complex.

All binding assays would be parallel assays, one of which contains the binding  
10 components alone and one of which contains the added candidate substance composition. One would perform each assay under conditions, and for a period of time, effective to allow the formation of protein:protein complexes or protein:DNA complexes, and one would then separate the bound complexes from any unbound protein and/or DNA and measure the amount of the complexes. An increase in the amount of any bound complex formed in the presence of the  
15 candidate substance would be indicative of a candidate substance capable of promoting BARD1 or other BRCA1 binding protein binding to BRCA1, or BARD1 or other BRCA1 binding protein-BRCA1 complex binding to DNA.

In such binding assays, the amount of the bound complex may be measured, after the  
20 removal of unbound species, by detecting a label, such as a radioactive or enzymatic label, which has been incorporated into the original wild-type, polymorphic or mutant BARD1, other BRCA1 binding protein or BRCA1 protein composition or even in a DNA segment. Alternatively, one could detect the protein portion of the complex by means of an antibody directed against the protein, such as those disclosed herein.

Preferred binding assays are those in which either the BARD1 or other BRCA1 binding  
25 protein or the BRCA1 protein is bound to a solid support and contacted with the other component to allow complex formation. Unbound protein components are then separated from the bound complexes by washing and the amount of the remaining bound complex is quantitated  
30 by detecting the label or with antibodies. Such binding assays form the basis of filter-binding and microtiter plate-type assays and can be performed in a semi-automated manner to enable analysis of a large number of candidate substances in a short period of time. Electrophoretic

methods of DNA binding, such as gel-shift assays, could also be employed to separate unbound protein or DNA from bound protein:DNA complexes.

5 Virtually any candidate substance may be analyzed by these methods, including compounds which may interact with BRCA1 or wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein, and also substances such as enzymes which may act by physically altering one of the structures present. Of course, any compound isolated from natural sources such as plants, animals or even marine, forest or soil samples, may be assayed, as may any synthetic chemical or recombinant protein.

10

Another potential method for stimulating BRCA1 activity is to prepare a wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein composition and to modify the protein composition in a manner effective to increase binding. The binding assays would be performed in parallel, similar to those described above, allowing the native and modified  
15 wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein binding to be compared. In addition to site specific mutagenesis, phosphatase and kinase enzymes may be tested, as may other agents, including proteases and chemical agents, could be employed to modify the BRCA1 binding properties of wild-type, polymorphic, mutant BARD1 or other BRCA1 binding proteins.

20

Cellular assays also are available for screening candidate substances to identify those capable of stimulating wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein and/or BRCA1-mediated transcription and gene expression. In these assays, the increased expression of any natural or heterologous gene under the control of a functional BRCA1 and  
25 wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein may be employed as a measure of stimulatory activity, although the use of reporter genes is preferred. A reporter gene is a gene that confers on its recombinant host cell a readily detectable phenotype that emerges only under specific conditions.

30

Reporter genes are genes which encode a polypeptide not otherwise produced by the host cell which is detectable by analysis of the cell culture, *e.g.*, by fluorometric, radioisotopic or spectrophotometric analysis of the cell culture. Exemplary enzymes include luciferases,

transferases, esterases, phosphatases, proteases (tissue plasminogen activator or urokinase), and other enzymes capable of being detected by their physical presence or functional activity. A reporter gene often used is chloramphenicol acetyltransferase (CAT) which may be employed with a radiolabeled substrate, or luciferase, which is measured fluorometrically.

5

Another class of reporter genes which confer detectable characteristics on a host cell are those which encode polypeptides, generally enzymes, which render their transformants resistant against toxins, *e.g.*, the *neo* gene which protects host cells against toxic levels of the antibiotic G418, and genes encoding dihydrofolate reductase, which confers resistance to methotrexate.

10 Other genes of potential for use in screening assays are those capable of transforming hosts to express unique cell surface antigens, *e.g.*, viral *env* proteins such as HIV gp120 or herpes gD, which are readily detectable by immunoassays.

The transcriptional promotion process which, in its entirety, leads to enhanced

15 transcription is termed "activation." The mechanism by which a successful candidate substance acts is not material since the objective is to promote wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein and/or BRCA1-mediated gene expression, or even, to promote gene expression in the presence of mutants, by whatever means will function to do so.

20 To create an appropriate vector or plasmid for use in such assays one would ligate the BRCA1 and wild-type, polymorphic, mutant BARD1 or other BRCA1 binding protein promoter and any necessary response elements to a DNA segment encoding the reporter gene by conventional methods. The relevant promoter sequences may be obtained by *in vitro* synthesis or recovered from genomic DNA and should be ligated upstream of the start codon of the

25 reporter gene. An AT-rich TATA box region should also be employed and should be located between the sequence and the reporter gene start codon. The region 3' to the coding sequence for the reporter gene will ideally contain a transcription termination and polyadenylation site. The promoter and reporter gene may be inserted into a replicable vector and transfected into a cloning host such as *E. coli*, the host cultured and the replicated vector recovered in order to

30 prepare sufficient quantities of the construction for later transfection into a suitable eukaryotic host.

Host cells for use in the screening assays of the present invention will generally be mammalian cells, and are preferably cell lines which may be used in connection with transient transfection studies. Cell lines should be relatively easy to grow in large scale culture. Also, they should contain as little native background as possible considering the nature of the reporter polypeptide. Examples include the Hep G2, VERO, HeLa, human embryonic kidney, 293, CHO, W138, BHK, COS-7, and MDCK cell lines, with monkey CV-1 cells being particularly preferred.

The screening assay typically is conducted by growing recombinant host cells in the presence and absence of candidate substances and determining the amount or the activity of the reporter gene. To assay for candidate substances capable of exerting their effects in the presence of mutated BARD1 or other BRCA1-binding gene products, one would make serial molar proportions of such gene products that alter expression. One would ideally measure the reporter signal level after an incubation period that is sufficient to demonstrate mutant-mediated repression of signal expression in controls incubated solely with mutants. Cells containing varying proportions of candidate substances would then be evaluated for signal activation in comparison to the suppressed levels. Candidates that demonstrate dose related enhancement of reporter gene transcription or expression are then selected for further evaluation as clinical therapeutic agents.

## VI. Diagnostics

As with the therapeutic methods of the present invention, the diagnostic methods are based upon the weight of evidence of the importance of *BARD1* and other genes identified herein, which encodes proteins that associate with BRCA1 *in vivo*. *BARD1* is co-expressed with BRCA1 in all breast and ovarian carcinoma lines tested. It is important to note that the BARD1/BRCA1 interaction is disrupted by tumorigenic amino acid substitutions in BRCA1, indicating that the formation of a stable complex between these proteins is likely to be an essential aspect of BRCA1-mediated tumor suppression. In this light, *BARD1* and the other genes encoding BRCA1-binding proteins are likely to be the target of oncogenic mutations in familial or sporadic breast cancer.

The diagnostic methods of the present invention generally involve determining either the type or the amount of a wild-type, polymorphic or mutant BARD1 or a BRCA1 binding protein present within a biological sample from a patient suspected of having breast, ovarian or another cancer. Irrespective of the actual role of BARD1 and the other BRCA1 binding proteins, it will be understood that the detection of a mutant is likely to be diagnostic of cancer and that the detection of altered amounts of BARD1 or one or more of the additional BRCA1 binding proteins, either at the mRNA or protein level, is also likely to have diagnostic implications, particularly where there is a reasonably significant difference in amounts.

The finding of a decreased amount of wild-type, polymorphic or mutant BARD1 or other BRCA1 binding protein in one, or preferably more, cancer patients, in comparison to the amount within a sample from a normal subject, will be indicative of BARD1 or one or more of the other BRCA1 binding proteins as a tumor suppressor. Following which, cancer in others would be similarly diagnosed by detecting a decreased amount of BARD1 or other BRCA1 binding protein in a sample. The finding of an increased amount of BARD1 or other BRCA1 binding protein in one, or preferably more, cancer patients, in comparison to the amount within a sample from a normal subject, will be indicative of BARD1 or one or more of the other genes encoding a BRCA1 binding proteins as an oncogene. Following which, cancer in others would be similarly diagnosed by detecting an increased amount of BARD1 or other gene encoding a BRCA1 binding protein in a sample.

The type or amount of a wild-type or mutant BARD1 or a BRCA1 binding protein present within a biological sample, such as a blood or tissue sample, may be determined by means of a molecular biological assay to determine the level of a nucleic acid that encodes such a BARD1 or BRCA1 binding protein, or by means of an immunoassay to determine the level of the polypeptide itself.

Any of the foregoing nucleic acid detection methods or immunodetection methods may be employed as a diagnostic methods in the context of the present invention.



## VII. Therapeutics

As stated above, the mechanism by which BRCA1 inhibits tumor formation is not yet completely understood. Most of the BRCA1 alleles that segregate with breast cancer susceptibility have frameshift or nonsense mutations that cause premature termination of protein synthesis, a relatively gross defect that provides fewer clues about the function of BRCA1 polypeptides.

In some families, however, the predisposing lesion of BRCA1 has been ascribed to a single amino acid substitution, such as the C61G and C64G mutations that occur within the RING domain. It is reasonable to propose that these mutations are oncogenic, at least in part, because they prevent the *in vivo* association of BRCA1 and BARD1 or other BRCA1 binding proteins. This suggests that the heteromeric BARD1/BRCA1 or other BRCA1 binding protein/BRCA1 complex has an active role in tumor suppression. This provides for two further aspects of the present invention.

First, the biochemical function of this protein complex can now be determined given that the present invention provides methods for obtaining sufficient amounts of the complex. The interaction between BARD1 and BRCA1 should situate their respective RING domains in close physical apposition. As such, the two domains could cooperatively perform certain functions, such as sequence-specific DNA recognition or association with other protein ligands. DNA recognition by the BARD1/BRCA1 complex is reasonable, especially since many transcription factors are known to bind DNA as obligate heterodimers (Landschulz *et al.*, 1988; Murre *et al.*, 1989). DNA recognition by complexes between BRCA1 and other BRCA1 binding proteins, even those that do not contain a RING motif, is also reasonable.

Second, upon confirmation of the active role of the heteromeric BARD1/BRCA1 or other BRCA1 binding protein/BRCA1 complex in tumor suppression, the present invention will provide cancer therapy by provision of the appropriate wild-type gene. The therapeutic methods are based upon the weight of evidence of the importance of *BARD1*, which encodes a protein that associates with BRCA1 *in vivo*, and is co-expressed with BRCA1 in all breast and ovarian carcinoma lines tested. Moreover, the *BARD1* gene product shares homology with the two most

highly conserved domains of BRCA1, both of which are common sites for germline mutations that segregate with breast cancer susceptibility. Finally, the BARD1/BRCA1 interaction is disrupted by tumorigenic amino acid substitutions in BRCA1, indicating that the formation of a stable complex between these proteins is likely to be an essential aspect of BRCA1-mediated tumor suppression.

In these aspects of the present invention, wild-type BARD1, or one of the genes encoding one of the other BRCA1-binding proteins disclosed herein, is provided to an animal with cancer, or breast, ovarian or uterine cancer, in the same manner that other tumor suppressors are provided, following identification of a cell type that lacks the tumor suppressor or that has an aberrant tumor suppressor. For example, the provision of BARD1, or one of the genes encoding one of the other BRCA1-binding proteins disclosed herein, can be considered to be analogous to the provision of p53.

Alternatively, should BARD1, or the gene encoding one of the other BRCA1 binding proteins, prove to be an oncogene, as may be established by the wild-type protein binding and reducing the activity of tumor suppressor proteins, then inhibition of BARD1, or the gene encoding one of the other BRCA1 binding proteins, would be adopted as a therapeutic strategy. This situation would be similar to that of MDM2, which binds and inhibits the tumor suppressor function of p53. Inhibitors would be any molecule that reduces the activity or amounts of BARD1 or a gene encoding one of the other BRCA1 binding proteins, including antisense, ribozymes and the like, as well as small molecule inhibitors.

### 1. Gene Therapy

The general approach to the tumor suppressor aspect of the present invention is to provide a cell with a wild-type or polymorphic BARD1 or a BRCA1 binding protein, thereby permitting the proper regulatory activity of the proteins to take effect. While it is conceivable that the protein may be delivered directly, a preferred embodiment involves providing a nucleic acid encoding a BARD1 or a BRCA1 binding protein to the cell. Following this provision, the polypeptide is synthesized by the transcriptional and translational machinery of the cell, as well as any that may be provided by the expression construct. In providing antisense, ribozymes and other inhibitors,

the preferred mode is also to provide a nucleic acid encoding the construct to the cell. All such approaches are herein encompassed within the term "gene therapy".

5 In various embodiments of the invention, DNA is delivered to a cell as an expression construct. Several non-viral methods for the transfer of expression constructs into cultured mammalian cells also are contemplated by the present invention. These include calcium phosphate precipitation, DEAE-dextran, electroporation, direct microinjection, DNA-loaded liposomes and lipofectamine-DNA complexes, cell sonication, gene bombardment using high velocity microprojectiles, and receptor-mediated transfection. Some of these techniques may be  
10 successfully adapted for *in vivo* or *ex vivo* use, as discussed below.

In another embodiment of the invention, the expression construct may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This  
15 is particularly applicable for transfer *in vitro*, but it may be applied to *in vivo* use as well.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter  
20 cells without killing them. Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force. The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

25 In a further embodiment of the invention, the expression construct may be entrapped in a liposome, as discussed below. Also contemplated are lipofectamine-DNA complexes. Liposome-mediated nucleic acid delivery and expression of foreign DNA *in vitro* has been very successful. Wong *et al.* (1980) demonstrated the feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa and hepatoma cells. In certain embodiments of the  
30 invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA. In other embodiments, the liposome may be complexed or employed in conjunction with

nuclear non-histone chromosomal proteins (HMG-1). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In other embodiments, the delivery vehicle may comprise a ligand and a liposome. Where a bacterial promoter is employed in the DNA construct, it also will be desirable to include within the liposome an appropriate bacterial polymerase.

The ability of certain viruses to enter cells via receptor-mediated endocytosis and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells. Preferred gene therapy vectors of the present invention will generally be viral vectors.

Retroviruses have promise as gene delivery vectors due to their ability to integrate their genes into the host genome, transferring a large amount of foreign genetic material, infecting a broad spectrum of species and cell types and of being packaged in special cell-lines (Miller, 1992).

Other viruses, such as adenovirus, herpes simplex viruses (HSV), cytomegalovirus (CMV), and adeno-associated virus (AAV), such as those described by U.S. Patent 5,139,941, incorporated herein by reference, may also be engineered to serve as vectors for gene transfer. Although some viruses that can accept foreign genetic material are limited in the number of nucleotides they can accommodate and in the range of cells they infect, these viruses have been demonstrated to successfully effect gene expression. However, adenoviruses do not integrate their genetic material into the host genome and therefore do not require host replication for gene expression, making them ideally suited for rapid, efficient, heterologous gene expression. Techniques for preparing replication-defective infective viruses are well known in the art.

In certain further embodiments, the gene therapy vector will be HSV. A factor that makes HSV an attractive vector is the size and organization of the genome. Because HSV is large, incorporation of multiple genes or expression cassettes is less problematic than in other smaller viral systems. In addition, the availability of different viral control sequences with varying performance (temporal, strength, *etc.*) makes it possible to control expression to a greater extent than in other systems. It also is an advantage that the virus has relatively few

spliced messages, further easing genetic manipulations. HSV also is relatively easy to manipulate and can be grown to high titers. Thus, delivery is less of a problem, both in terms of volumes needed to attain sufficient MOI and in a lessened need for repeat dosings.

5       Of course, in using viral delivery systems, one will desire to purify the virion sufficiently to render it essentially free of undesirable contaminants, such as defective interfering viral particles or endotoxins and other pyrogens such that it will not cause any untoward reactions in the cell, animal or individual receiving the vector construct. A preferred means of purifying the vector involves the use of buoyant density gradients, such as cesium chloride gradient  
10       centrifugation.

Gene delivery using second generation retroviral vectors has been reported. Kasahara *et al.* (1994) prepared an engineered variant of the Moloney murine leukemia virus, that normally infects only mouse cells, and modified an envelope protein so that the virus  
15       specifically bound to, and infected, human cells bearing the erythropoietin (EPO) receptor. This was achieved by inserting a portion of the EPO sequence into an envelope protein to create a chimeric protein with a new binding specificity.

## 2.       Antisense

20

In an alternative embodiment, the BARD1 or BRCA1 binding protein nucleic acids employed may actually encode antisense constructs that hybridize, under intracellular conditions, to BARD1 or BRCA1 binding protein nucleic acids. The term "antisense construct" is intended to refer to nucleic acids, preferably oligonucleotides, that are complementary to the  
25       base sequences of a target DNA or RNA. Antisense oligonucleotides, when introduced into a target cell, specifically bind to their target nucleic acid and interfere with transcription, RNA processing, transport, translation and/or stability.

Antisense constructs may be designed to bind to the promoter and other control regions,  
30       exons, introns or even exon-intron boundaries of a gene. Antisense RNA constructs, or DNA encoding such antisense RNA's, may be employed to inhibit gene transcription or translation or both within a host cell, either *in vitro* or *in vivo*, such as within a host animal, including a human

subject. Nucleic acid sequences which comprise "complementary nucleotides" are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, that the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T), in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

As used herein, the terms "complementary" means nucleic acid sequences that are substantially complementary over their entire length and have very few base mismatches. For example, nucleic acid sequences of fifteen bases in length may be termed complementary when they have a complementary nucleotide at thirteen or fourteen positions with only a single mismatch. Naturally, nucleic acid sequences which are "completely complementary" will be nucleic acid sequences which are entirely complementary throughout their entire length and have no base mismatches.

Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology, but also contains a non-homologous region (e.g., a ribozyme) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

While all or part of the BARD1 or BRCA1 binding protein gene sequence may be employed in the context of antisense construction, short oligonucleotides are easier to make and increase *in vivo* accessibility. However, both binding affinity and sequence specificity of an antisense oligonucleotide to its complementary target increases with increasing length. One can readily determine whether a given antisense nucleic acid is effective at targeting of the corresponding host cell gene simply by testing the constructs *in vitro* to determine whether the function of the endogenous gene is affected or whether the expression of related genes having complementary sequences is affected.

In certain embodiments, one may wish to employ antisense constructs which include other elements, for example, those which include C-5 propyne pyrimidines. Oligonucleotides

which contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression.

### VIII. Pharmaceutical Compositions

#### A. Pharmaceutically Acceptable Carriers

Aqueous compositions of the present invention comprise an effective amount of the BARD1 or other BRCA1 binding agent, such as a BARD1 or other BRCA1 binding protein, peptide, epitopic core region, inhibitor, or such like, dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. Aqueous compositions of gene therapy vectors expressing any of the foregoing are also contemplated. The phrases "pharmaceutically or pharmacologically acceptable" refer to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, or a human, as appropriate.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

For human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

The biological material should be extensively dialyzed to remove undesired small molecular weight molecules and/or lyophilized for more ready formulation into a desired vehicle, where appropriate. The active compounds will then generally be formulated for parenteral administration, *e.g.*, formulated for injection via the intravenous, intramuscular, subcutaneous, intralesional, or even intraperitoneal routes. The preparation of an aqueous composition that contains a BARD1 or other BRCA1 binding agent as an active component or

ingredient will be known to those of skill in the art in light of the present disclosure. Typically, such compositions can be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for using to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and the preparations can also be emulsified.

5

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

10

Solutions of the active compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

15

A BARD1 or other BRCA1 binding protein, peptide, agonist or antagonist of the present invention can be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

20

25

30

The carrier can also be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of



microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

In terms of using peptide therapeutics as active ingredients, the technology of U.S. Patents 4,608,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, each incorporated herein by reference, may be used.

The preparation of more, or highly, concentrated solutions for direct injection is also contemplated, where the use of DMSO as solvent is envisioned to result in extremely rapid penetration, delivering high concentrations of the active agents to a small tumor area.

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above, but drug release capsules and the like can also be employed.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous,

intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

The active BARD1- or other BRCA1 binding protein-derived peptides or agents may be formulated within a therapeutic mixture to comprise about 0.0001 to 1.0 milligrams, or about 0.001 to 0.1 milligrams, or about 0.1 to 1.0 or even about 10 milligrams per dose or so. Multiple doses can also be administered.

In addition to the compounds formulated for parenteral administration, such as intravenous or intramuscular injection, other pharmaceutically acceptable forms include, *e.g.*, tablets or other solids for oral administration; liposomal formulations; time release capsules; and any other form currently used, including cremes.

One may also use nasal solutions or sprays, aerosols or inhalants in the present invention. Nasal solutions are usually aqueous solutions designed to be administered to the nasal passages in drops or sprays. Nasal solutions are prepared so that they are similar in many respects to nasal secretions, so that normal ciliary action is maintained. Thus, the aqueous nasal solutions usually are isotonic and slightly buffered to maintain a pH of 5.5 to 6.5.

In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, and appropriate drug stabilizers, if required, may be included in the formulation. Various commercial nasal preparations are known and include, for example, antibiotics and antihistamines and are used for asthma prophylaxis.

Additional formulations which are suitable for other modes of administration include vaginal suppositories and pessaries. A rectal pessary or suppository may also be used.

Suppositories are solid dosage forms of various weights and shapes, usually medicated, for insertion into the rectum, vagina or the urethra. After insertion, suppositories soften, melt or dissolve in the cavity fluids.

5

In general, for suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%.

10

Vaginal suppositories or pessaries are usually globular or oviform and weighing about 5 g each. Vaginal medications are available in a variety of physical forms, *e.g.*, creams, gels or liquids, which depart from the classical concept of suppositories. Vaginal tablets, however, do meet the definition, and represent convenience both of administration and manufacture.

15

Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders.

20

In certain defined embodiments, oral pharmaceutical compositions will comprise an inert diluent or assimilable edible carrier, or they may be enclosed in hard or soft shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, 25 elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 75% of the weight of the unit, or preferably between 25-60%. The amount of active compounds in such therapeutically useful compositions is such that a suitable dosage will be obtained.

30

The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a

disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compounds sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor.

It will naturally be understood that suppositories, for example, will not generally be contemplated for use in treating breast cancer. However, in the event that the proteins, peptides or other agents of the invention, or those identified by the screening methods of the present invention, are confirmed as being useful in connection with other forms of cancer, then other routes of administration and pharmaceutical compositions will be more relevant. As such, suppositories may be used in connection with colon cancer, inhalants with lung cancer and such like.

#### **B. Liposomes and Nanocapsules**

In certain embodiments, the use of liposomes and/or nanoparticles is contemplated for the introduction of wild-type, polymorphic or mutant BARD1 or other BRCA1 binding protein peptides or agents, or gene therapy vectors, including both wild-type and antisense vectors, into host cells. The formation and use of liposomes is generally known to those of skill in the art, and is also described below.

Nanocapsules can generally entrap compounds in a stable and reproducible way. To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1  $\mu\text{m}$ ) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention, and such particles may be easily made.

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4  $\mu$ m. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous solution in the core.

The following information may also be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

Liposomes interact with cells via four different mechanisms: Endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or *vice versa*, without any association of the liposome contents. Varying the liposome formulation can alter which mechanism is operative, although more than one may operate at the same time.

### C. Kits

Therapeutic kits of the present invention are kits comprising a wild-type, polymorphic or mutant BARD1 and/or other BRCA1 binding protein, peptide, inhibitor, gene, vector or other BARD1 or BRCA1 binding protein effector. Such kits will generally contain, in suitable

container means, a pharmaceutically acceptable formulation of a BARD1 or BRCA1 binding protein, peptide, domain, inhibitor, or a gene or vector expressing any of the foregoing in a pharmaceutically acceptable formulation, optionally comprising other anti-cancer agents. The kit may have a single container means, or it may have distinct container means for each compound.

5

When the components of the kit are provided in one or more liquid solutions, the liquid solution is an aqueous solution, with a sterile aqueous solution being particularly preferred. The BARD1 and BRCA1 binding protein compositions may also be formulated into a syringeable composition. In which case, the container means may itself be a syringe, pipette, or other such like apparatus, from which the formulation may be applied to an infected area of the body, injected into an animal, or even applied to and mixed with the other components of the kit.

10

However, the components of the kit may be provided as dried powder(s). When reagents or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. It is envisioned that the solvent may also be provided in another container means.

15

The container means will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the BARD1 or BRCA1 binding protein or gene or inhibitory formulation are placed, preferably, suitably allocated. Where a second anti-cancer therapeutic is provided, the kit will also generally contain a second vial or other container into which this agent may be placed. The kits may also comprise a second/third container means for containing a sterile, pharmaceutically acceptable buffer or other diluent.

20

The kits of the present invention will also typically include a means for containing the vials in close confinement for commercial sale, such as, e.g., injection or blow-molded plastic containers into which the desired vials are retained.

25

Irrespective of the number or type of containers, the kits of the invention may also comprise, or be packaged with, an instrument for assisting with the injection/administration or placement of the ultimate BARD1 or BRCA1 binding protein or gene composition within the

30

body of an animal. Such an instrument may be a syringe, pipette, forceps, or any such medically approved delivery vehicle.

5       The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate  
10       that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

## EXAMPLE I

### Methods

#### 1. Two-hybrid screening in yeast

A cDNA fragment encoding the amino-terminal 304 residues of human BRCA1 was obtained by RT-PCR™ amplification of HeLa cell RNA with flanking oligonucleotide primers:

TTACCATGGATTTATCTGCTCTTCGCGTT (SEQ ID NO:4); and

20       AAAAGTCGACTAGAAATTCAGCCTTTTCTACATTCATTC (SEQ ID NO:5).

After digestion with *Nco*I and *Sal*I endonucleases, the amplified fragment was inserted into the corresponding sites of the pAS1-CYH2 vector (Harper *et al.*, 1993). The resultant plasmid (BR304/pAS1-CYH2) was then used to transform yeast cells of the Y190 reporter strain  
25       (Trp<sup>-</sup> Leu<sup>-</sup> His<sup>-</sup>, LacZ<sup>-</sup>). Trp<sup>+</sup> prototrophs were evaluated for expression of the DBD-BR304 hybrid polypeptide (containing the GAL4 DNA-binding domain fused to the amino-terminal 304 residues of BRCA1) by immunoblotting with 12CA5, a monoclonal antibody that recognizes the influenza hemagglutinin epitope incorporated into the expressed reading frame of pAS1-CYH2 (Chien *et al.*, 1991).

30       These cells were then transfected with a cDNA library of human B cell transcripts in the pACT two-hybrid expression vector (Clontech), and approximately 11 million Trp<sup>+</sup>Leu<sup>+</sup> transformants were plated on a Trp/Leu/His dropout medium containing 40 mM 3-aminotriazole

(Durfee *et al.*, 1993). The positive clones ( $\text{His}^+\text{LacZ}^+$ ) were cured of the BR304/pAS1-CYH2 plasmid by growth on Leu dropout plates containing 10 mg/ml cycloheximide (Harper *et al.*, 1993).

5 Each of the cured clones was then subjected to a two-hybrid mating assay for protein-protein interactions with the DBD-BR304 hybrid and DBD hybrids containing sequences of two irrelevant proteins (mouse p53 and human TAL1). The cDNAs that displayed a BRCA1-specific pattern of interaction in the mating assay were excised from the library plasmid (pACT), inserted into pAS1-CYH2, and tested for BRCA1-specific interaction in a reciprocal two-hybrid  
10 mating assay with BR304/pACTII, an expression vector that encodes a hybrid protein (TAD-BR304) containing the transactivation domain of GAL4 fused to the amino-terminal 304 residues of BRCA1.

Three of the DBD-X hybrid proteins, including the DBD-STAT3 hybrid and two DBD-  
15 X hybrids encoded by novel cDNA sequences, could not be tested in the reciprocal yeast two-hybrid assay because they were self-activating; that is, they were able to induce expression of the LacZ reporter construct in the absence of the TAD-BR304 hybrid.

## 2. Two-hybrid analysis in mammalian cells

20 Candidate cDNAs that showed a BRCA1-specific pattern of interaction in yeast were also subjected to two-hybrid analysis in mammalian cells (Dang *et al.*, 1991; Hsu *et al.*, 1994). For this purpose, each cDNA was inserted into the multiple cloning site of pVP-HA2 or pVP-FLAG, mammalian vectors designed for the expression of hybrid polypeptides that contain the transactivation domain of the herpesvirus VP16 protein. In addition, sequences encoding  
25 BRCA1 residues 1-304 were inserted into pM1, a mammalian vector used for expression of hybrid proteins containing the DNA-binding domain of GAL4 (Sadowski *et al.*, 1992). Embryonal kidney 293 cells were then co-transfected with an expression vector encoding the candidate VP16 hybrid polypeptide (3.0 mg), an expression vector encoding the GAL4-BR304 hybrid (BR304/pM1) (3.0 mg), a GAL4-responsive reporter gene (G5LUC) (1.0 mg), and the  
30 pSV- $\beta$ -galactosidase control plasmid (1.5 mg).



Expression vectors for mammalian two-hybrid analyses of the BARD1/BRCA1 interaction (FIG. 4A, FIG. 4B and FIG. 5A) were constructed by inserting defined cDNA segments into pVP-HA2, pVP-FLAG, or pCMV-GAL4; the latter, which is a derivative of the pCMV5 (Andersson *et al.*, 1989) and pM2 (Andersson *et al.*, 1989) vectors, contains a sequence encoding the FLAG epitope appended to the 3' end of the GAL4 reading frame.

### 3. Antibody production

The bacterial expression vector encoding GST-BRΔ304, a glutathione S-transferase fusion protein containing residues 183-304 of human BRCA1, was generated by inserting a BRCA1 cDNA fragment into the *NcoI/HindIII* sites of pGEX-KG. The fusion protein was then expressed in *E. coli*, isolated to homogeneity by affinity chromatography on glutathione-agarose, and injected into rabbits according to a standard immunization protocol. Similarly, the BARD1-specific antiserum was generated by immunizing rabbits with a purified GST-fusion protein containing BARD1 residues 141-388. The TAL1-specific antiserum (#1080) has been described (Hsu *et al.*, 1994).

### 4. Co-immunoprecipitation analysis

The TAL1 expression plasmid (TAL1/pCMV4) has been described (Hsu *et al.*, 1994). The expression plasmid for HA-BR304 was constructed in two steps: First, the cDNA fragment encoding residues 1-304 of human BRCA1 was inserted into the *NcoI/SalI* sites of pVP-HA2, a vector used for expression of VP16-fusion proteins in mammalian cell. Second, the BRCA1 coding sequences were excised from pVP-HA2, along with vector sequences encoding the influenza hemagglutinin (HA) epitope, and inserted into the *NotI/HindIII* sites of pCMV-*Not*, a derivative of the pCMV4 expression vector (Andersson *et al.*, 1989).

The vectors encoding FLAG-DE12 and FLAG-B202 were also prepared in two steps: thus, the appropriate cDNA fragments were inserted into pVP-FLAG, and the cDNA fragments were then excised from pVP-FLAG, together with vector sequences encoding the FLAG epitope, and inserted into the *NotI/HindIII* sites of pCMV-*Not*.

For co-immunoprecipitation analysis, approximately  $25 \times 10^5$  embryonal 293 kidney cells were seeded onto each 100 mm plate and cultured in 10 ml of growth medium (low glucose

DMEM supplemented with 2 mM glutamine, 100 mg/ml penicillin G, 100 mg/ml streptomycin, and 10% fetal calf serum). After 24 hours the adherent cells were treated with the calcium phosphate transfection system according to the manufacturer's instructions (Gibco/BRL). Each 100 mm culture was transfected with 3.75 mg of the pSV- $\beta$ -galactosidase control plasmid (Promega) and 7.5 mg of each expression vector; where necessary 7.5 mg of the parental pCMV4 vector was added to provide a constant DNA mass (18.75 mg) for transfection of each culture.

Two days after transfection, cell lysates were prepared in 1 ml of "low-salt NP40 buffer" (10 mM HEPES pH 7.6, 250 mM NaCl, 0.1% Nonidet P-40, 5 mM EDTA) containing protease inhibitors (0.1 mg/ml aprotinin, 1 mg/ml leupeptin, 1 mg/ml pepstatin, and 1 mM PMSF), and 4 ml of immune or pre-immune rabbit antiserum were added to each lysate. After rocking at 4°C for 1 hr, 50 ml of staphylococcal protein A-Sepharose beads (20% slurry; Pharmacia) were added to each lysate and the mixture was rocked at 4°C for an additional hour. The beads were then pelleted by brief centrifugation and washed two times in "high-salt NP40 buffer" (10 mM HEPES pH 7.6, 1.0 M NaCl, 0.1% Nonidet P-40, 5 mM EDTA) with protease inhibitors and two times in low-salt NP40 buffer with protease inhibitors.

Finally, the beads were resuspended in "loading buffer" (100 mM Tris-HCl pH 6.8, 2% SDS, 0.2% bromophenol blue, 20% glycerol, and 5%  $\beta$ -mercaptoethanol), boiled for 10 minutes, and pelleted by centrifugation. The supernatant was then fractionated by electrophoresis on a SDS-15% polyacrylamide gel, and the fractionated polypeptides were electroblotted onto Hybond-ECL nitrocellulose for Western analysis by enhanced chemiluminescence (Amersham) with the FLAG-specific M5 monoclonal antibody (Eastman Kodak).

## 5. *In vitro* assays of protein-protein interaction

Expression plasmids encoding the full-length BARD1 and BRCA1 polypeptides were generated by inserting their respective cDNA fragments into pSP6-FLAG, a derivative of the pSPUTK vector (Stratagene) that includes coding sequences for an amino-terminal tag containing the FLAG epitope (MADYKDDDKS; SEQ ID NO:3) (Hopp *et al.*, 1988). The BARD1/pSP6-FLAG and BRCA1/pSP6-FLAG plasmids were then used as templates for

*in vitro* synthesis of radiolabeled BARD1 and BRCA1 polypeptides, respectively, in rabbit reticulocyte lysates (Promega) containing [ $S^{35}$ ]methionine (DuPont NEN).

Expression plasmids encoding GST-fusion proteins were generated by inserting the appropriate cDNA fragments into the pGEX or pGEX-KG vectors (Smith and Johnson, 1988; Guan and Dixon, 1991). The GST fusion proteins were expressed in *E. coli*, purified by affinity chromatography on glutathione-agarose beads, and retained as a 50% slurry in "buffer C" (20 mM Hepes pH 7.6, 100 mM KCl, 1 mM EDTA, 1 mM dithiothreitol and 20% glycerol) with protease inhibitors (Smith and Johnson, 1988).

The loaded beads were then used directly in binding assays with radiolabeled full-length BARD1 polypeptides. Thus, for each binding reaction, a 10 ml aliquot of the BARD1-programmed reticulocyte lysate was mixed with 100 ml of glutathione-agarose beads (loaded with 10 mg of the GST-fusion protein) and 890 ml of "low-salt binding buffer" (50 mM Hepes pH 7.6, 250 mM NaCl, 0.5% Nonidet P-40, 5 mM EDTA, 0.1% bovine serum albumin, 0.5 mM dithiothreitol, 0.005% SDS, and protease inhibitors). Following a 1 hr incubation at room temperature, the beads were washed twice with low-salt binding buffer, twice with high-salt binding buffer (containing 1M NaCl), and twice again with low-salt binding buffer. Finally, the beads were boiled for 10 minutes in 80 ml of loading buffer, and 40 ml of the supernatant was fractionated by electrophoresis on a SDS-10% polyacrylamide gel.

*In vitro* co-immunoprecipitation was performed by mixing 50 ml of rabbit reticulocyte lysate containing radiolabeled full-length FLAG-BRCA1 with 50 ml of reticulocyte lysate containing unlabeled full-length FLAG-BARD1 or with 50 ml of an uncharged reticulocyte lysate. Each mixture was incubated at 37°C for 30 minutes in the presence of protease inhibitors. Equivalent aliquots of the mixtures (19 ml) were then diluted into 960 ml of low-salt NP40 buffer and immunoprecipitated at 4°C for 1 hour with 20 ml of staphylococcal protein A-Sepharose beads (50% slurry; Pharmacia) and 1 ml of the indicated antiserum. The beads were then pelleted by brief centrifugation and washed four times in low-salt NP40 buffer. Finally, the beads were resuspended in loading buffer, boiled for 10 minutes, and pelleted by centrifugation. The supernatant was then fractionated by electrophoresis on a SDS-6% polyacrylamide gel.

## 6. Expression studies

Cytoplasmic RNA was isolated from breast and ovarian cancer cell lines by a combination of NP-40 lysis and mechanical disruption before the addition of lysates to guanidinium isothiocyanate (Sambrook *et al.*, 1989). Total RNA was subjected to electrophoresis and blotted as described (Sambrook *et al.*, 1989). The probe for *BARD1* was purified cDNA insert from the B202 or B230 clones. The 18S probe was obtained from the ATCC (#77242). Probes were labeled by random hexanucleotide extension with [<sup>32</sup>P]dCTP (Amersham).

Northern blots were hybridized at 42°C in 50% formamide solution containing dextran sulfate (Oncor) for 48 hours and subjected to a final wash in 0.5X SSC, 0.1% SDS at 65°C. Hybridization signals were quantitated after overnight exposure to a PhosphorImager (PI) screen using Imagequant software (Molecular Dynamics). Blots were then exposed to X-ray film; 18S was exposed for 20 minutes to the PI screen and for 2 hours to X-ray film.

## 7. Chromosomal localization of *BARD1*

The location of *BARD1* was determined by PCR™ amplification of a panel of monochromosomal hybrid DNAs obtained from the Coriell Institute; using the human *BARD1* primers:

B202L, AACAGTACAATGACTGGGCTC; SEQ ID NO:6; and  
B202R, TCAGCGCTTCTGCACACAGT; SEQ ID NO:7.

The location of *BARD1* was further refined by mapping in the Genebridge panel of DNAs from whole genome radiation hybrids.

## 8. Clinical Specimens

Tumor tissue, matched normal tissue and blood specimens were obtained as part of protocols approved by the University of Texas Southwestern Medical Center Human Subjects Review Board, St. Paul's Medical Center, Medical City of Dallas and The Southern division of the Cooperative Human Tissue Network. The breast cancers were primarily infiltrating ductal carcinomas. The ovarian carcinomas were of mixed histology, although the majority were

papillary serous carcinomas. The following breast and ovarian cancer cell lines were obtained from the American Type Culture Collection: MCF-7, ZR75-1, BT-483, BT-20, T-47D, BT-474, 2008, OVCAR3, CAOV-3, BG-1 and 2774. The ovarian cancer line PE04 was obtained from Dr. Simon Langdon (Medical Oncology Unit, Western General Hospital, Edinburgh, Scotland).

5 Tumors were immediately frozen in liquid nitrogen and stored at -70°C prior to RNA extraction. Buffy coat was prepared from blood. In some cases DNA was prepared from paraffin-embedded tissue. DNA, RNA and cDNA was prepared by standard procedures (Sambrook, *et al.*, 1989).

#### 9. Genomic structure of BARD1

10 A human genomic library was first screened by hybridization with fragments of BARD1 cDNA (Example IV, below). Eleven hybridizing lambda clones were identified and subjected to nucleotide sequence analysis with oligonucleotide primers derived from BARD1 cDNA sequence and shown in Table 4 (see Example X below).

15 YACs lying between D2S143 and D2S295 (The location of BARD1) were identified by accessing the Whitehead data-base. YACs containing BARD1 were identified on the basis that they generated the correctly sized PCR amplification products with primers for exons for which genomic sequence was available as a result of sequencing lambda clones. These YACs were sized on pulsed-field gels and isolated as described elsewhere (Gemmill *et al.*, 1996) and YACs  
20 810d12 and 964g6 were then subcloned into the cosmid vector sCos-1 as described (Clines *et al.*, 1997). Hybridization of this library of approximately 5,000 cosmids with probes derived from amplification with BARD1 cDNA primers described in Table 4 (B230-F/FAS, B230-FF/FFAS, B230-WS/WAS) resulted in the identification of eleven positively hybridizing cosmids. The same primers were used to sequence two of these cosmid DNAs, generating  
25 exon/intron boundary sequences for this region, for which lambda clones were not available.

#### 10. Mutational screening for BARD1 alterations

cDNA was derived from tumor, matched normal tissue or cell lines. Genomic DNA was obtained from tumor tissue, matched normal tissue, cell lines, blood, and paraffin embedded  
30 tissue. SSCP was performed as described elsewhere (Orita *et al.*, 1989; Orita *et al.*, 1989) with oligonucleotide primers for BARD1 with cDNA or genomic DNA as shown in Tables 4 and 5 (see Examples X and XI below).

Briefly, PCR<sup>TM</sup> of tumor or blood DNA/cDNA was performed in 20 $\mu$ l volumes containing 100 ng cDNA or genomic DNA template; 1 $\times$  PCR buffer (Perkin Elmer, Foster City, CA); 200  $\mu$ M each dATP, dGTP, dCTP, dTTP; 10 pmoles each primer (GIBCO BRL, Grand Island, NY); 0.3 $\mu$ Ci <sup>32</sup>P-dCTP (Amersham, Arlington Heights, IL); 0.5U Taq DNA polymerase (Perkin Elmer, Foster City, CA). PCR<sup>TM</sup> conditions were 30 cycles of 94°C for 30 seconds; 55°C (or as specified for annealing temperatures in Tables 4 and 5) for 30 seconds; 72°C for 30 seconds. A final extension reaction at 72°C was performed for 1 minute.

Amplified samples were diluted 1:10 in formamide buffer (98% formamide, 10 mM EDTA, pH, 8.0, 0.05% bromophenol blue, 0.05% xylene cyanol), denatured at 95°C for 5 min. then cooled rapidly to 4°C. For each sample, 4  $\mu$ l was loaded onto an SSCP gel and run at 8W (constant power) for 8-16 hours in 0.6 $\times$  TBE at room temperature. Gels contained 0.5  $\times$  MDE (AT Biochem), 0.6 $\times$  TBE, 240  $\mu$ l 10% ammonium persulphate, 24  $\mu$ l TEMED. Duplicate gels were prepared with a supplement of 10% glycerol. Gels were subjected to autoradiography with or without being dried. Film was exposed for 12-24h. with an intensifying screen.

#### 11. DNA Sequencing of BARD1 Variants Identified by SSCP

Variant bands were excised from the SSCP gel after alignment with the autoradiograph and purified with Qiaquick Gel Extraction kit (Qiagen, Santa Clarita, CA, Cat # 28706). DNA was resuspended in 20  $\mu$ l H<sub>2</sub>O and 5  $\mu$ l was treated with 10 units exonuclease I and 2 units shrimp alkaline phosphatase at 37°C for 15 min. Following inactivation of this reaction with heat (80°C for 15 min.), the DNA template was subjected to cycle sequencing with Thermosequenase (Amersham Life Science, Arlington Heights, IL) and  $\alpha$ -<sup>32</sup>P-ddNTPs. Sequencing reactions were electrophoresed in 8% acrylamide/bis gels with 1 $\times$  glycerol tolerant gel buffer at 70W constant power for 2 hours. Gels were dried and subjected to autoradiography.

#### 12. Fish Mapping of BARD1

The cytogenetic location of BARD1 was obtained with fluorescence *in situ* hybridization (FISH) of normal human metaphase chromosome spreads with phage DNA pooled from three of the lambda clones (R12, R5 and R35). One microgram of DNA was labeled with biotin using

DOP-PCR (Telenius *et al.*, 1992) and subjected to FISH analysis as described elsewhere (Trask, 1997; Wise *et al.*, 1997).

### 13. Preparation of Normal Breast cDNA Library

5 Total RNA was isolated from normal breast tissue obtained during reduction mammoplasty surgery and flash frozen. Approximately 1 gram pieces of tissue (containing fat, epithelium, stroma and normal vessels, *etc.*) was ground in 8 ml 4 M guanidinium isothiocyanate solution by a virtishear blender. The lysate was layered over 3 ml of a 5.7 M CsCl solution and centrifuged at 32K for 18 hours at 20°C in a Beckman SW41T rotor.

10 Total RNA pellets were resuspended phenol/chloroform extracted and reprecipitated. RNA pellets were resuspended in DEPC H<sub>2</sub>O and concentration measured by spectrophotometry at OD<sub>260</sub>.

15 Aliquots of total RNA (approximately 10 µg) were electrophoresed on 1.2% agarose formaldehyde denaturing gels to assess intact status of the 28S and 18S ribosomal RNAs.

Total RNA from 3 separate patients was pooled (nB 63 10.6%, nB 52 45.6%, nB 62 43.9%). The total RNA samples were not treated with DNase I before isolation of poly A<sup>+</sup> RNA. Poly A<sup>+</sup> RNA was isolated by two passages over oligo dT Dynabeads, with regeneration of the beads in between isolation rounds.

20 Approximately 5 µg of poly A RNA was used to prepare the cDNA library. The library was prepared in the pACT two-hybrid expression vector (Clontech, Palo Alto, CA), and then used in the yeast two hybrid screening method as detailed in section 1 above.

### EXAMPLE II

#### Yeast two-hybrid screening with the amino-terminal sequences of BRCA1

30 A cDNA sequence encoding the amino-terminal 304 residues of BRCA1 was amplified by RT-PCR™ and inserted into the pAS1-CYH2 expression vector (Harper *et al.*, 1993). The resultant plasmid (BR304/pAS1-CYH2) encodes a hybrid protein containing the DNA-binding domain of GAL4 fused to BRCA1 residues 1-304. Yeast cells of the Y190 reporter strain

(Harper *et al.*, 1993) were then transformed in succession with the BR304/pAS-CYH2 plasmid and with an expression library of human B cell cDNAs fused to sequences encoding the GAL4 transactivation domain (Durfec *et al.*, 1993).

5 By screening approximately 11 million library transformants, the inventors isolated 312 clones that co-activate the GAL4-responsive *HIS3* and *lacZ* reporter genes of Y190. Forty-six of the isolates were found to interact specifically with BRCA1 in a yeast two-hybrid mating assay that employed two irrelevant proteins (mouse p53 and human TAL1) as negative controls (Harper *et al.*, 1993). Nucleotide sequence analysis revealed that the 46 isolates represent  
10 twenty-six independent cDNA clones derived from sixteen distinct mRNAs. The candidate BRCA1-associated proteins encoded by these cDNAs are comprised of eleven novel polypeptides and five known proteins; the latter include TAFII70/80 (Genbank accession nos. L25444 and U31659), filamin (X53416), STAT3/APRF (L29277), UNPH (U20657), and a human homolog of the yeast GCN5 gene product (U57317).

15 The eleven novel polypeptides are BARD1 (SEQ ID NO:2); and the genes encoding the TCL52 (SEQ ID NO:9), TCL163 (SEQ ID NO:10), B223 (SEQ ID NO:11), B115 (SEQ ID NO:12), BAP28 (SEQ ID NO:13), B48 (SEQ ID NO:14), B258 (SEQ ID NO:15), BAP152 (SEQ ID NO:16), B123 (SEQ ID NO:17) and B268 (SEQ ID NO:18) polypeptides.

20 Each of the candidate proteins was also tested in a reciprocal yeast two-hybrid study in which residues 1-304 of BRCA1 were expressed as a fusion protein with the GAL4 transactivation domain (TAD-BR304) and the candidate cDNA sequence was expressed as a fusion with the GAL4 DNA-binding domain (DBD-X). Three of the DBD-X hybrid  
25 polypeptides were capable of activating the reporter genes in the absence of TAD-BR304, obviating their analysis in the reciprocal two-hybrid assay. However, each of the other thirteen DBD-X hybrids registered as positive in this assay; that is, reporter gene activation occurred in the presence of the TAD-BR304 hybrid but not in the presence of control hybrids, such as TAD-TAL1 and TAD-SV40 large T antigen.



### EXAMPLE III

#### Protein-protein interactions in mammalian cells

Additional tests were conducted to determine whether any of the candidate proteins  
5 interact with BRCA1 in mammalian cells. Therefore, a mammalian expression plasmid was  
prepared which encodes GAL4-BR304, a protein containing the DNA-binding domain of GAL4  
fused to BRCA1 residues 1-304. In addition, expression vectors that encode each of the  
candidate BRCA1-associated proteins as hybrids with the VP16 transactivation domain were  
also prepared.

10 The mammalian version of the two-hybrid assay was then performed by transfecting  
human 293 kidney cells with a GAL4-responsive reporter gene (G5LUC) and pairwise  
combinations of the appropriate expression vectors (Dang *et al.*, 1991; Hsu *et al.*, 1994).  
Transcription of the reporter gene was evaluated by measuring the luciferase activity of lysates  
15 prepared from the transfected cells.

As illustrated in FIG. 1, expression of the GAL4-BR304 hybrid did not induce  
significant luciferase activity in transfected 293 cells (see lane 1). Likewise, expression of  
20 VP16-B202, a VP16-hybrid that contains sequences from one of the candidate BRCA1-  
associated proteins, also failed to activate transcription of the G5LUC reporter gene (lane 10).  
However, co-expression of GAL4-BR304 and VP16-B202 generated a large increase in  
luciferase activity to levels more than 30-fold greater than those found with either hybrid alone  
(lane 9). This suggests that the BRCA1 and B202 moieties of the hybrid polypeptides interact  
stably with one another in mammalian cells. In contrast, pairwise expression of GAL4-BR304  
25 with each of the other six VP16-fusion proteins did not yield a measurable increase in luciferase  
activity (lanes 3, 5, 7, 11, 13, and 15).

To date, fifteen of the sixteen candidate BRCA1-associated proteins have been tested for  
interaction with BRCA1 in the mammalian two-hybrid system; all of these proteins, with the  
30 exception of B202, failed to associate with BRCA1 in the mammalian assay.

Co-immunoprecipitation studies were carried out to confirm that the BRCA1 and B202 polypeptides interact in mammalian cells. Therefore, an expression plasmid was prepared that encodes HA-BR304, a polypeptide containing the amino-terminal tag:

MAYPYDVPDYASLRS, SEQ ID NO:8, appended to residues 1-304 of BRCA1.

A plasmid was also constructed for expression of FLAG-B202, a polypeptide that includes an amino-terminal tag with the FLAG epitope, MADYKDDDDKS; SEQ ID NO:3 (Hopp *et al.*, 1988), and 177 residues encoded by B202.

Human 293 cells were co-transfected with different combinations of these expression plasmids and, as controls, plasmids that encode two helix-loop-helix transcription factors (E12 or TAL1) that are known to form stable heterodimers *in vivo* (Hsu *et al.*, 1994). Two days after transfection the cells were lysed under mild conditions. Aliquots of each lysate were immunoprecipitated with either a rabbit antiserum raised against residues 183-304 of human BRCA1, the corresponding pre-immune serum, or a TAL1-specific antiserum.

To determine whether the FLAG-B202 polypeptide was co-immunoprecipitated with HA-BR304, the precipitates were fractionated by SDS-PAGE, and the presence of FLAG-B202 was determined by immunoblotting with a monoclonal antibody (M5; Eastman Kodak) that recognizes the FLAG epitope. FLAG-B202 was co-immunoprecipitated with the BRCA1-specific antiserum, but not with the corresponding pre-immune serum or with an antiserum specific for TAL1. Moreover, co-immunoprecipitation of FLAG-B202 was clearly dependent on the presence of HA-BR304 since it was not observed using lysates of cells expressing FLAG-B202 alone. Therefore, a specific *in vivo* association between B202 and BRCA1 can be demonstrated in mammalian cells by two independent procedures, the two-hybrid assay and co-immunoprecipitation analysis.

#### EXAMPLE IV

##### The BRCA1-associated RING-domain (BARD1) protein

The B202 clone, which contains a cDNA insert of ~1.0 kilobasepairs, represents five of the 46 isolates obtained in the yeast two-hybrid screen. An independent isolate (B230)

contained a distinct but overlapping insert of 2.5 kilobasepairs. The composite cDNA sequence of 2,531 bp (SEQ ID NO:1) derived from B202 and B230 includes a large open reading frame with at least two potential initiator codons and encodes a protein with the sequence of SEQ ID NO:2. Translation from the first two initiation methionines (residues M1 and M26) would generate polypeptides of 777 and 752 amino acids, respectively. Residue 153 of SEQ ID NO:2 is denoted with the letter "X" to reflect a difference between the sequence of B202 and B230; the corresponding triplet in these cDNAs encodes a lysine (AAA) or glutamic acid (GAA) residue, respectively. Significantly, a cysteine-rich domain (residues 46-90) that matches the consensus sequence of the RING motif of BRCA1 and the PML1 and BMI-1 oncoproteins is found near the amino-termini of these polypeptides.

The BRCA1-associated RING domain protein (designated BARD1) also contains a centrally-located sequence comprised of three tandem ankyrin repeats (residues 427-525), a 33-amino acid motif found in a variety of different regulatory proteins (Bork, 1993). In addition, when the BLAST algorithm was used to screen protein databases with the remaining BARD1 sequences on the carboxy-terminal side of the ankyrin repeats (Altschul *et al.*, 1990), a significant homology with BRCA1 (and only BRCA1) was uncovered.

Moreover, the homologous region of BRCA1 corresponds to the phylogenetically-conserved sequence that lies near its carboxy-terminus (Sharan *et al.*, 1995). Recently, Koonin *et al.* showed that this sequence bears a weak but significant homology with the carboxy-terminal regions of the mammalian 53BP1 protein, the yeast RAD9 gene product, and two putative proteins encoded by uncharacterized cDNA clones (Koonin *et al.*, 1996). The homologous sequences are comprised of two tandem copies of the BRCA1 carboxy-terminal domain (the "BRCT domain"), a newly recognized amino acid motif of unknown function (Koonin *et al.*, 1996).

Although homology with 53BP1 was not detected in a conventional BLAST search of existing protein databases with the BARD1 sequence, the similarity of their carboxy-terminal regions becomes apparent when each is independently aligned with the BRCT domains of BRCA1. Within each of these proteins the levels of sequence identity between the first and second copies of the BRCT domain are modest; nevertheless, the homology between the tandem

copies is illustrated when the core motifs of each, which consist of a relatively well-conserved stretch of 38 amino acids, are aligned with one another (Koonin *et al.*, 1996). Thus, BARD1 and BRCA1 belong to a small family of proteins that harbor BRCT domains at their carboxy-termini. Within this family BARD1 and BRCA1 are especially related in that they also possess  
5 an amino-terminal RING motif (FIG. 2).

#### EXAMPLE V

##### In vitro analysis of the BARD1/BRCA1 interaction

10 To examine the binding properties of BARD1 and BRCA1 *in vitro*, cDNA sequences encoding the full-length polypeptides were inserted into the pSPUTK expression vector (Stratagene) along with a short amino-terminal tag containing the FLAG epitope (MADYKDDDDKS; SEQ ID NO:3). The resultant plasmids (BARD1/pSP6-FLAG and BRCA1/pSP6-FLAG, respectively) were then used as templates for coupled *in vitro*  
15 transcription/translation in rabbit reticulocyte lysates.

Radiolabeled full-length BARD1 polypeptides were generated by *in vitro* translation in a rabbit reticulocyte lysate. An aliquot (0.2 ml) of the lysate was fractionated by electrophoresis on a SDS-10% polyacrylamide gel. Additional aliquots (10 ml) were incubated with purified  
20 GST-fusion proteins loaded onto glutathione-agarose beads. The washed beads were boiled in 80 ml of loading buffer, and equivalent aliquots of the eluants (40 ml) were fractionated by electrophoresis. The binding reactions were conducted with parental GST, GST-BR304, GST-TAL1, GST-E47, GST-ATF4, GST-BR184, or GST-BRD304.

25 Translation of BARD1/pSP6-FLAG in the presence of [<sup>35</sup>S]methionine generated a radiolabeled BARD1 polypeptide of ~97 kilodaltons. Equivalent aliquots of the radiolabeled protein were then mixed with purified glutathione S-transferase (GST) or with purified GST-fusion proteins containing various segments of BRCA1 or segments of the TAL1, E2A, or ATF4 transcription factors. After a short incubation, the GST proteins of each mixture were absorbed  
30 to glutathione-agarose beads.

The radiolabeled BARD1 polypeptide was retained on the beads by the GST-BR304 fusion protein (which contains BRCA1 residues 1-304), but not by the parental GST polypeptide

or by GST fusion proteins containing irrelevant sequences from TAL1, E2A, or ATF4. Moreover, *in vitro* binding of BARD1 was observed with the GST-BR184 fusion protein (which contains BRCA1 residues 1-184) but not with the GST-D304 polypeptide (which contains BRCA1 residues 183-304). These results suggest that BARD1 and BRCA1 polypeptides  
5 interact directly to form a stable protein complex *in vitro*, and that the interaction is mediated by sequences within the amino-terminal 184 residues of BRCA1.

Although in most of these assays the BARD1/BRCA1 interaction was evaluated using segments of one or both polypeptides, the ability of the full-length proteins to associate with one  
10 another was also examined. For this purpose, full-length BRCA1 was generated by *in vitro* translation in a rabbit reticulocyte lysate containing [<sup>35</sup>S]methionine, while full-length BARD1 was produced by *in vitro* translation in an unlabeled reticulocyte lysate. The radiolabeled BRCA1 lysate was then incubated with the unlabeled BARD1 lysate or with an uncharged reticulocyte lysate, and equivalent aliquots of the mixture were subjected to  
15 immunoprecipitation with antisera specific for BRCA1, BARD1, or TAL1, or with preimmune serum as a control, and fractionated on a SDS-6% polyacrylamide gel.

As now expected, the BRCA1-specific antiserum, but not the corresponding pre-immune serum, immunoprecipitated full-length BRCA1 from the mixture along with a series of smaller  
20 degradation products. Significantly, the BRCA1 polypeptides were also co-immunoprecipitated from the mixture with a BARD1-specific antiserum but not with an antiserum raised against TAL1. Co-immunoprecipitation of BRCA1 with the BARD1-specific antiserum was clearly dependent on the presence of BARD1, since it was not observed when radiolabeled BRCA1 was mixed with an unlabeled reticulocyte lysate that did not contain *in vitro*-translated BARD1  
25 polypeptides. These results indicate that the full-length BARD1 and BRCA1 polypeptides can interact to form a stable protein complex.

## EXAMPLE VI

### Expression and chromosomal localization of the *BARD1* gene

30 Northern hybridization revealed two major BARD1 transcripts (5.9 and 4.4 kilobases) in all the breast and ovarian cancer cell lines tested (ZR-75, T-47D, BT-483, Ovarcar-3, Caov3, 2774, 2008).

The chromosomal location of *BARD1* was determined by PCR™ amplification of a panel of monochromosomal hybrid DNAs with primers specific for *BARD1* (B202L and B202R; SEQ ID NO:6 and SEQ ID NO:7, respectively). A single human-specific band of 230 basepairs was seen in the hybrid containing a single human chromosome 2. The location of *BARD1* was further refined by mapping in the Genebridge panel of DNAs from whole genome radiation hybrids. This analysis placed *BARD1* in the distal region of human chromosome 2q, 3.56 cR distal to D2S143 (lod >3.0) and flanked by D2S295 distally.

## EXAMPLE VII

### The interacting regions of BARD1

The sequences of BARD1 that interact with BRCA1 should be located within the shared segment encoded by both B202 and B230 (amino acid residues 8-311) - the two independent BARD1 cDNA clones obtained in the yeast two-hybrid screen (see FIG. 2). These sequences were further localized by mammalian two-hybrid studies in which smaller segments of BARD1 (FIG. 2) were expressed as fusion proteins with the VP16 transactivation domain.

As illustrated in FIG. 3, VP16-fusion proteins containing segments NB (residues 26-202) and NE (residues 26-142), both of which encompass the RING domain of BARD1 (residues 46-90), readily activated the GAL4-responsive reporter gene when expressed in the presence of GAL4-BR304, the GAL4-fusion protein containing residues 1-304 of BRCA1 (lanes 3 and 5). BRCA1 association was also observed in reciprocal two-hybrid assays in which the NB and NE segments of BARD1 were expressed as GAL4-fusion proteins and tested for interaction with VP16-BR304. Therefore, the interaction with BRCA1 is mediated by sequences in the vicinity of the BARD1 RING domain.

## EXAMPLE VIII

### The interacting regions of BRCA1

The *in vitro* binding studies showed that the interacting sequences of BRCA1 reside within its amino-terminal 184 residues. These sequences were further localized by mammalian two-hybrid analysis with VP16-NE, a hybrid polypeptide containing the VP16 transactivation

domain fused to the NE segment of BARD1 (residues 26-142). VP16-NE was tested for interaction with a panel of GAL4-hybrid proteins containing different amino-terminal segments of BRCA1.

5 As shown in FIG. 4A, the BR147 (residues 1-147) and BR101 (residues 1-101) segments, both of which encompass the RING motif of BRCA1 (residues 20-68), retain the ability to interact with BARD1 (lanes 3 and 5). However, BARD1-association was not achieved with a smaller segment that also includes the intact RING domain (BR71, residues 1-71) (FIG. 4A, lane 7), despite the fact that the GAL4-BR71 hybrid protein was expressed at levels  
10 comparable to those of GAL4-BR147 and GAL4-BR101, as judged by western analysis with the M5 anti-FLAG monoclonal antibody.

The same result was obtained from a reciprocal two-hybrid study in which GAL4-BR304 was tested for binding with VP16-hybrids containing different segments of BRCA1 (FIG. 4B).  
15 Thus, although association between BARD1 and BRCA1 is mediated by sequences in the immediate vicinity of their respective RING motifs, the RING domain of BRCA1 is not by itself sufficient to mediate the interaction.

#### EXAMPLE IX

##### Tumorigenic missense mutations of BRCA1

20 The tumorigenic missense mutations of BRCA1 were analyzed in regard to their effect on the BARD1/BRCA1 interaction. Since the C61G and C64G mutations eliminate conserved zinc-binding cysteines from the RING motif of BRCA1, the inventors sought to determine the  
25 effect of these mutations on BARD1/BRCA1 association. Therefore, C61G and C64G substitutions were incorporated into the BR304 segment of BRCA1 by site-directed mutagenesis of the corresponding cDNA fragment. Expression plasmids were then constructed to encode GAL4-BR304 hybrid polypeptides that contain either the C61G (GAL4-BR304-C61G) or C64G (GAL4-BR304-C64G) lesion.

30

As illustrated in FIG. 5A, the wild-type GAL4-BR304 hybrid (lane 3), but not its mutant derivatives (lanes 5 and 7) interacted with BARD1 in the mammalian two-hybrid assay, despite

the fact that all three versions of the GAL4-BR304 polypeptide were expressed at comparable levels, as judged by western analysis with the M5 anti-FLAG monoclonal antibody.

The effect of the missense mutations on BARD1/BRCA1 association was also evaluated by co-immunoprecipitation studies of mammalian cell lysates (FIG. 5B). Thus, 293 cells were co-transfected with expression plasmids encoding FLAG-B202 (described above) and either a wild-type or mutant derivative of FLAG-BR304, a BR304 polypeptide with an amino-terminal tag containing the FLAG epitope (MADYKDDDDKS; SEQ ID NO:3). Two days later the cells were lysed and aliquots of each lysate were immunoprecipitated with either the BRCA1-specific antiserum or the corresponding pre-immune serum.

To determine whether FLAG-B202 polypeptides were co-immunoprecipitated with FLAG-BR304, the immunoprecipitates were fractionated by SDS-PAGE, and the presence of FLAG-B202 was determined by immunoblotting with the M5 anti-FLAG monoclonal antibody. FLAG-B202 was co-immunoprecipitated with the BRCA1-specific antiserum when expressed in the presence of wild-type FLAG-BR304 (FIG. 5B; lane 2). In contrast, however, co-immunoprecipitation did not occur when FLAG-B202 was expressed with FLAG-BR304 derivatives containing either the C61G or C64G substitutions (lanes 4 and 6).

Together, the mammalian two-hybrid and co-immunoprecipitation studies demonstrate that the C61G and C64G mutations prevent formation of an *in vivo* protein complex between BRCA1 and BARD1.

## EXAMPLE X

### Genomic structure of BARD1

To obtain the genomic DNA encoding BARD1, lambda phage and cosmid libraries of human genomic or YAC DNA (YACs 810d12 and 964g6) were first screened by hybridization with fragments of BARD1 cDNA (Example IV, above). Eleven hybridizing lambda clones and two hybridizing BAC clones were subjected to nucleotide sequence analysis with oligonucleotide primers derived from BARD1 cDNA sequence (Table 4, below). This analysis resulted in nine large contigs of genomic sequence (SEQ ID NO:122, containing exon 1 and 5'



untranslated region (UTR), which likely contains the BARD1 promoter; SEQ ID NO:123, containing exon 2 and exon 3; SEQ ID NO:124, containing exon 4; SEQ ID NO:125, containing exon 5; SEQ ID NO:126, containing exon 6; SEQ ID NO:127, containing exon 7; SEQ ID NO:128, containing exon 8; SEQ ID NO:129, containing exon 9; and SEQ ID NO:130, containing exon 10 and exon 11, plus 3' UTR; from the 5' end of the gene to the 3' end of the gene, respectively), which revealed that the BARD1 coding sequences are derived from eleven exons distributed over at least 65 kilobases of genomic DNA.

The chromosomal origin of BARD1 was then established by fluorescence in-situ hybridization (FISH) of normal human chromosomes with subclones containing BARD1 genomic sequences. FISH analysis localized BARD1 to bands 2q34-35, consistent with the BARD1 mapping data obtained previously with the Genbridge panel of whole genome radiation hybrid DNAs (Example VI, above).

**TABLE 4**  
**PCR™ Primers for the Amplification of BARD1 Sequences from cDNA Template**

Bp	Forward Primer (5'>3')	SEQ ID #	Reverse Primer (5'>3')	SEQ ID #	PCR™ Prod. Size	Ann. Temp.
44	GCGAGGAGCCTTTCATCCGA	57	CGAGCGGGCGGACTGT	58	154	59
149	ATGGAACCGGATGGTCGGGT	59	TCITCAAGTCITGTATCCAGGC	60	205	59
145	CGCCATGGAACCAATACA	61	TCITCAAGTCITGTATCCAGGC	60	209	57
340	GCCTGGATACAAGACTTGAAG	62	TTGTAGACGTCCTCCTGAACC	63	306	57
551	AAAGCTTCAGTGC AAACCCA	64	TCCAGATCTTGCAGAAAGCC	65	132	53
638	CAGATGTTTCTGAGAGGGCT	66	ATTCTCTTTGGAGTCAAATTC	67	138	55
734	GAGGCAGAAAAGAAAGATGGT	68	AGGAGCCACTTGTCTAGTAAG	69	136	55
855	ATGGTGAAATAGACTTACTAGC	70	GCAGACCITCTCAGGAGTC	71	149	55
946	AAGAGCAGGAATGAAGTAGTG	72	CTCCACTGGTGCTCAGAATG	73	163	55
1103	AGTGGAGATTTTGTAAAGCAA	74	AGGTGGTGTAGGTGGTGAA	75	159	51
1250	GGTACACCAACCTTCTACATT	76	GTCTCTCTCTCTATGATTTCTT	77	113	53
1311	CAATGAAGCTGTTGCCCAA	78	GTCTTTAACAATTTGGATCACT	79	137	51
1427	AGTGATCCAAATGTTAAAGAC	80	CCCATCTTGGCTGCATC	81	162	51
1550	CAAATGACTCACCACTTCAC	82	ATCGACAGGCCGACAGACC	83	120	55
1661	CCTGTCGATTATACAGATGAT	84	AACATGAGTTACTGTACTGTC	85	234	57
1862	TATACTGAGTTTGACAGTACAG	86	CATACITTTCTTCGTAGACATG	87	146	55
1976	B230-G: TGGGTAAAGCATGTCTACGA	88	TCAGCGCTTCTGCACACAGT	7	126	55
2093	B230-H: GGATGCTACTTCTATTGTG	89	GAGTCACGTCACCTGTCTG	90	134	51
2179	B230-TS: CCTCAGTAGAAAGCCCAAGC	91	GCCCCTGCCGAACCTCTC	92	154	57
2215	B230-US: GAGAGGGTTCGGCAGGGC	93	TTCAATTTCAAATGTTTCATCTGTT	94	124	57

**EXAMPLE XI****BARD1 mutation screening**

The inventors used SSCP (Orita *et al.*, 1989a; Orita *et al.*, 1989b) to screen genomic  
5 DNA or cDNA from 48 breast tumors, 58 ovarian tumors, 60 uterine cancers (primarily  
endometrial), six breast cancer lines and six ovarian cancer lines and germline DNA or  
lymphoblastoid-derived cDNA from 67 breast/ovarian cancer patients with no observed  
alterations in BRCA1 or BRCA2 for genetic alterations in BARD1. SSCP was performed as  
described elsewhere (Orita *et al.*, 1989; Orita *et al.*, 1989) with oligonucleotide primers for  
10 BARD1 with cDNA or genomic DNA as shown in Table 4 (Example X above) and Table 5  
(below). Variant bands were excised from the SSCP gel, subjected to a second round of  
amplification and sequenced.

TABLE 5

## PCR™ Primers for the Amplification of BARD1 Sequences from Genomic DNA Template

Exon	Forward Primer (5'>3')	SEQ ID #	Reverse Primer (5'>3')	SEQ ID #	PCR™ Size (bp)	Ann.Temp. (°C)
I	GCGAGGAGCCCTTTCATCCGA	57	CGAGCGCGCGCGACTGT	58	154	59
I	ACAGTCGCGCGCGCTCGA	95	CAGAAACTGTGCGACCCGTG	96	107	59
II	AGATGTTTATCTAACAATGACTC	97	AGTTGTACTATATACATCAAACC	98	146	55
III	ATTCTGCTGAATGGGTGCTT	99	TAACTAAGAGAGATAGGGATAG	100	226	55
IVa	GGAGCTCCATGTGGGAGCAA	101	AACATCTGCAGGAGGACTTGG	102	270	59
IVb	CAGATGTTTCTGAGAGGGCT	66	ATTCTCTTTGGAGTCAAAATTC	67	138	55
IVc	GAGGCAGAAAAGAGATGGT	68	AGGAGCCACTTGCTAGTAAG	69	136	55
IVd	ATGGTGAAATAGACTTACTAGC	70	GCAGACCTTCTCAGGAGTC	71	149	55
IVe	AAGAGCAGGAATGAAGTAGTG	72	CTCCACTGGTGCTCAGAATG	73	163	55
IVf	AGTGGAGATTTTGTTAAGCAA	74	AGGTGGTGTAGTGGTGAA	75	159	51
IVg	GGTACACCCACCTTCTACATT	76	TCTGAGATGGTATTTTCAGAGT	103	170	53
V	TGCTTTTAAATTTCATTTTGTTC	104	AAGAACTGTAAACACAGAAAGA	105	163	55
VI	TGCTCTTTCTTATCACTTCTTTC	106	CTTGACTCAAGAATATAGGTCC	107	278	57
VII	TTGAGTCGAGTCACACATTTGA	116	CTATTATGTTCCCTTTCATAACCA	117	233	55
VIII	TAATGTCTTTTGCTAGTCTGTAA	118	GGTAGTTCCTCCAAAAGGATCA	119	264	55
IX	GAGTTATAAGAAGCAGGCCAA	120	ATTCTTAATTCTCTCAAAATCCAA	121	199	55
X	TAGTGCTCACITGATACITTAGT	108	CATAATAAGAAACAATGAAAGTTGT	109	187	55
XIa	TTGATCTGCCCTTTAACAAATG	110	GCCCCCTGCCGGAACCCCTCTC	92	296	57
XIb	GAGAGGGTTCGGCAGGGC	93	TTCAATTTCAAATGTTTCATCTGTT	94	124	57

### A. BARD1 Mutations

When 58 ovarian tumors were analyzed, one (ov61) was found to harbor a missense mutation within BARD1 that resulted in a glutamine to histidine (CAG to CAC; Q564H; SEQ ID NO:32 (nucleic acid) and SEQ ID NO:33 (amino acid)) change between the ankyrin repeats and the BRCT domain (FIG. 6). This patient was a woman of African-American origin who was diagnosed at age 73 with a clear cell adenocarcinoma of the ovary (stage 3A) and a synchronous infiltrating lobular carcinoma of the breast. Only the mutant allele was detected in the ovarian tumor cDNA from this individual, indicating that the wild-type transcript was either expressed at undetectable levels or was completely absent. The absence of detectable wild-type fragments indicates that the ovarian carcinoma cells of the patient were devoid of normal BARD1 polypeptides. At the time of hysterectomy six years earlier this patient had been diagnosed with an incidental stage IA endometrial clear cell tumor. It is likely that these represent two separate primary tumors of the endometrium and ovary since the initial endometrial tumor was a small focus of carcinoma confined to an endometrial polyp.

Genomic DNA extracted from paraffin-embedded tissue obtained from the three primary tumors, as well as from benign uterine tissue, were examined from this patient. SSCP analysis identified the variant allele in all samples, including normal uterine tissue, indicating that this alteration was of germ-line origin. Moreover, the wild-type allele of BARD1 was absent from the genomic DNA of the ovarian tumor, explaining the loss of wild-type BARD1 transcripts. Both the wild-type and mutant alleles were detected in genomic DNA of both the endometrial and breast cancers; however, histological examination indicated that a significant proportion of normal tissue had infiltrated these tumor specimens. This contaminating normal tissue could have obscured the ability to detect loss of the wild-type allele in the breast and endometrial tumors. The high degree of infiltrating normal tissue also rendered microdissection of tumor tissue from these samples impossible.

The Q564H missense alteration was not seen in over 300 individuals examined (>600 chromosomes), suggesting that this alteration is not a polymorphism. Since this patient was African American, an additional 30 African individuals (60 chromosomes) were screened for this variant. The variant was not detected, indicating that this change is unlikely to be a polymorphism private to the African population. In light of the interaction of BARD1 with

BRCA1, and the observed loss of the wild-type BARD1 allele in the ovarian tumor, the germ-line missense alteration, Q564H, may have resulted in predisposition to endometrial, breast and ovarian cancer. Additionally, since the glutamine 564 residue is conserved in the mouse sequence, it is likely to be of some importance.

5 A second ovarian tumor (ov208) harbored a variant within the BRCT domain (FIG. 6). This tumor was obtained from a 16 year old Caucasian female and was diagnosed as a small cell carcinoma of the ovary with neuroendocrine features. The genetic alteration in this tumor resulted in an arginine to cysteine change at amino acid 658 (R658C; SEQ ID NO:36 (nucleic  
10 acid) and SEQ ID NO:37 (amino acid)). This alteration was only seen in one other sample; an endometrial adenocarcinoma obtained from a 67 year old woman (ut14). This change was not seen in any other DNAs examined (>600 chromosomes). The alteration in ov208 was determined to be of germ-line origin. In this ovarian tumor sample the wild-type allele was detected, but it is not known if this was derived from contaminating normal tissue present in this  
15 tumor sample, and therefore whether the wild-type allele had been lost from the tumor itself.

As a result of the Q564H finding, the Inventors became interested in the involvement of BARD1 in the development of uterine tumors and examined an additional ten for alterations. One had a serine to asparagine change at amino acid 761 in the BRCT domain (S761N; SEQ ID  
20 NO:34 (nucleic acid) and SEQ ID NO:35 (amino acid)). This alteration (S761N) occurs in the 3' end of the BRCT domain, and lies within the 30 amino acid core motif of BRCT domains adjacent to the invariant tryptophan residue. The wild-type allele was also detected in this tumor.

25 No mutations were seen in the germ-line DNA of the 67 breast/ovarian cancer patients. None of these had reported BRCA1/2 mutations, although none have been screened fully for such mutations. All these patients, except one, had a family history of cancer (43 breast/ovarian, 22 breast and 2 ovarian).

30 Alterations of BARD1 in sporadic breast and ovarian tumors appear to be a rare event. This observation is correlated with the fact that 2q, the location of BARD1, has not been reported to undergo significant LOH in breast/ovarian cancer. However, it is possible that

BARD1, like BRCA1 is involved in tumorigenesis through other mechanisms such as alterations in transcript level (Thompson *et al.*, 1995). The low frequency of genetic alterations in BARD1 in breast and ovarian tumors is similar to findings for BRCA1 and BRCA2. In the case of BRCA1, no genetic alterations have been detected in sporadic breast tumors. However, 10% of ovarian tumors harbor somatic mutations that result in protein truncations. In these tumors there is also loss of the wild-type allele (Hosking *et al.*, 1995; Merajver *et al.*, 1995).

In the case of BRCA2, four independent studies collectively identified two sporadic missense alterations and one somatic truncating mutation in 281 primary breast cancers and two somatic alterations in 185 ovarian carcinomas (Lancaster *et al.*, 1996; Miki *et al.*, 1996; Phelan *et al.*, 1996; Takahashi *et al.*, 1996; Teng *et al.*, 1996; Weber *et al.*, 1996). The alteration in one of the ovarian carcinomas was an "A" insertion in one poly(A) tract of the gene due to a mutation in the DNA mismatch repair gene hMSH2 (Takahashi *et al.*, 1996). The second ovarian carcinoma had a missense mutation of unknown significance.

Despite the rarity of the BARD1 alterations in tumors of the breast, ovary and endometrium, loss of its wild-type allele in the ovarian tumor ov61 provides evidence for a tumor-suppressor role (Haber and Harlow, 1997) for BARD1 in the prevention of these cancers. The BARD1 alteration in this tumor, Q564H, occurred between the BRCT domains and the ankyrin repeats. The function of the BRCT domains of BARD1 is unknown, although in the case of BRCA1 this region has been shown to have transactivational function (Chapman and Verma, 1996; Montiero *et al.*, 1996).

The homology of the BRCT domain with domains in proteins such as RAD9, XRCC1 and RAD4, which are involved in cell cycle checkpoint functions in response to DNA damage (Bork *et al.*, 1997; Callebaut and Mornon, 1997; Koonin *et al.*, 1996), and the recent finding that BRCA1 associates with another DNA repair protein, RAD51 (Scully *et al.*, 1997), suggests that it may be important in mediating repair of DNA damage. Together with BRCA1, BARD1 may be involved in cell cycle checkpoint control in response to DNA damage. The inventors have recently found further evidence for a common role for these two proteins by demonstrating that BRCA1 and BARD1 co-localize in nuclear dots in the S phase of the cell cycle (Example XIV below).

90% of germ-line alterations in BRCA1 and all germ-line alterations in BRCA2 that predispose to breast/ovarian cancer result in protein truncation (Shattuck-Eidens *et al.*, 1995; Stratton, 1996). However, in the case of p53, missense mutations are the most common alteration in human breast cancer as they are in other tumors. The recently isolated PTEN/MMAC1 gene, which is altered in Cowden disease (Liaw *et al.*, 1997) as well as in sporadic brain, prostate and kidney cancers (Li *et al.*, 1997; Steck *et al.*, 1997), has been reported to harbor both nonsense and missense mutations. These are predicted to disrupt the protein tyrosine/dual-specificity phosphatase domain of the PTEN/MMAC gene product.

#### **B. BARD1 Polymorphisms**

Seven polymorphic sites were detected within BARD1. A description of BARD1 polymorphic sites and variants is shown in FIG. 6 and described below.

One polymorphism was detected in the first exon, 5' to the region encoding the RING domain. This mutation is a proline to serine change at amino acid 24 (P24S; SEQ ID NO:20 (nucleic acid) and SEQ ID NO:21 (amino acid)).

A second polymorphism was detected as a result of sequencing two cDNA clones that differed at nucleotide 531. This mutation is a lysine (AAA) to glutamic acid (GAA) change at amino acid 153 (SEQ ID NO:22 (nucleic acid) and SEQ ID NO:23 (amino acid)).

Primers C/CAS amplify a region located between the RING domain and the first ankyrin repeat. Two polymorphisms (polymorphisms three and four) were seen within this region. The third polymorphism is a C to G transversion at nucleotide 1121, generating a silent polymorphism within a threonine codon (CCG to CGG; amino acid 351; SEQ ID NO:24 (nucleic acid) and SEQ ID NO:25 (amino acid)).

The fourth polymorphism was a deletion of seven amino acids (PLPECSS) between amino acids 358 and 364 (SEQ ID NO:26 (nucleic acid) and SEQ ID NO:27 (amino acid)). When individuals that were not selected because of a family history of breast/ovarian cancer, were examined, this deletion was seen in 2/68 individuals from the CEPH (Centre du



Polymorphisme Humain) but was not detected in 40 other Caucasian individuals ascertained in the United States. This deletion appeared to be in linkage disequilibrium with the "G" allele at nucleotide 1121. This deletion was only seen in 2/216 unrelated Caucasian chromosomes where there was no significant family history of breast/ovarian cancer, but was far more frequent in  
5 Africa as it was seen in 1/15 chromosomes. This accounts for its higher frequency in African-American women and in tumors from this population in general.

Interestingly, both the MCF7 cell line and the PEO4 ovarian cancer cell line harbored this deletion. In both these cell lines both alleles were expressed. MCF7 was developed from a  
10 pleural effusion of a 69 year old Caucasian woman with a malignant mammary adenocarcinoma (Soule *et al.*, 1973). PEO4 was developed from the peritoneal ascites of a Caucasian woman with an a poorly differentiated serous adenocarcinoma (Langdon *et al.*, 1988). An African-American woman who developed ovarian endometrioid adenocarcinoma at the age of 68 was homozygous for this deletion. However, since the frequency of this deletion is 0.067 in  
15 Africans, the frequency of homozygotes is 0.005 in African populations. The frequency of a homozygote in African-Americans would be expected to be lower than this, so that within the sample set of DNA samples from approximately 100 African-American individuals, detection of one homozygote is not an impossibility.

20 A fifth polymorphism was seen in the third ankyrin repeat, and resulted in a valine to methionine change at amino acid 507 (V507M; SEQ ID NO:28 (nucleic acid) and SEQ ID NO:29 (amino acid)).

A sixth polymorphism was located between the ankyrin repeats and the BRCT domain.  
25 This results in a cysteine to serine change at amino acid 557 as a result of a G to C transversion (C557S; SEQ ID NO:30 (nucleic acid) and SEQ ID NO:31 (amino acid)). This polymorphism was also seen in the BT474 breast cancer cell line (Lasfargues *et al.*, 1978).

A seventh polymorphism was located in the BRCT domain. This results in a serine to  
30 asparagine change at amino acid 761 (S761N; SEQ ID NO:38 (nucleic acid) and SEQ ID NO:39 (amino acid)). It is also possible that this alteration occurs at a much lower frequency that would be more indicative of a mutation than a polymorphism.

However, gene deletions do not necessarily account for disease or cancer susceptibility. For example, a polymorphic stop codon within the 3' end of the coding sequence of BRCA2 results in loss of the 93 most terminal amino acids (Lys3326ter) with as yet no described deleterious effect (Mazoyer *et al.*, 1996).

## EXAMPLE XII

### Other BRCA1-interacting Clones:

#### A. Clones Isolated From a Breast cDNA Library

Four additional genes which encode proteins that interact with BRCA1 were detected in the breast cDNA library using the yeast two-hybrid screening assay described in Example I above. The genes isolated were designated BE2 (SEQ ID NO:40 (nucleic acid) and SEQ ID NO:41 (amino acid)), BE14 (SEQ ID NO:42 (nucleic acid) and SEQ ID NO:43 (amino acid)), BE31 (SEQ ID NO:44 (nucleic acid) and SEQ ID NO:45 (amino acid)) and BE445 (SEQ ID NO:46 (nucleic acid) and SEQ ID NO:47 (amino acid)).

BE2 encodes a 1.25 kb transcript in spleen, prostate, testes, small intestine, colon, and ovary. An additional transcript of approximately 1.0 kb is also seen in testes. It is also transcribed in some breast/ovarian cancer lines (Table 6, below). BE14 encodes a 4.4 kb transcript in testes.

**TABLE 6****BE2 Expression in Breast and Ovarian Cancer Cell Lines**

Type	Cell Line Name	ATCC #	BE2 expression
Breast Cancer	BT-474	HTB 20	-
	BT-483	HTB 121	-
	MDA-MB-134 VI	HTB 23	-
	MDA0MB-361	HTB 27	+
	Ly-2		-
	MCF-7	HTB 22	-
	T-47D	HTB 133	-
	ZR-75-1	CRL 1500	-
	BT-20	HTB 19	-
	MDA-MB-231	HTB 26	+++
	MDA-MB-436	HTB 130	+
	MDA-MB-453	HTB 131	-
	MDA-MB-468	HTB 132	-
	MDA-MB-435S	HTB 129	+++
	SCC 38		+
	SCC 70		+
	BT-549	HTB 122	++
	SCC 202		-
	SCC 712		-
	SCC 1007		-
Ovarian Cancer	2008		-
	2774		-
	CaOv-3	HTB 75	+++
	OVCAR-3	HTB 161	-
	PA1	CRL 1572	+++
	PEO4		++
	SKOV-3	HTB 77	++
	SW626	HTB 78	+
	UCI 101		-
	UCI 109		-
	SCC 60		++
	SCC 1426		+
	SCC 1159		++

5 **B. Genomic Mapping of Additional BRCA1 Binding Clones**

The BE2 gene was mapped with gene-specific primers and genome-wide radiation hybrids to 11p15, the locale of a tumor suppressor gene for breast, ovarian and lung cancer

(Winqvist *et al.*, 1993). The possibility exists that this is the tumor suppressor gene that maps to this location.

The BE14 gene was mapped with gene-specific primers and genome-wide radiation hybrids to chromosome 3q. This gene encodes a 4.4 kb transcript that we have only seen in testis. Like BRCA1, BRCA2 and BARD1, this gene is transcribed in breast cancer cells that have been starved by treatment with charcoal-stripped fetal calf serum and then supplemented with estrogen (Example XIII below). This suggests that all these genes are estrogen responsive, or are induced after the cells have been signaled to proliferate by signals created as a result of estrogen binding the estrogen receptor. This may have implications relating to the therapeutic aspects of these genes.

The B123 gene has been localized to 17pter, the locale of a tumor suppressor gene for breast cancer (Cropp *et al.*, 1990; Lindblom *et al.*, 1993).

### EXAMPLE XIII

#### Estrogen Responsiveness of BRCA1, BRCA2 and BARD1

##### A. Methods

##### 1. Cell Culture

The previously characterized breast cancer cell lines BT-483 (Lasfargues *et al.*, 1978) and MCF-7 were obtained from the American Type Culture Collection (ATCC No. HTB121 and HTB22). BT-483 cells were routinely cultured in RPMI 1640 media containing phenol red, 2 mM glutamine and 1X antibiotic/antimycotic solution (Life Technologies, Gaithersburg, MD) supplemented with 20% fetal calf serum (FCS) (Life Technologies) and 10 µg/ml bovine insulin (Sigma, St. Louis, MO) in a humidified atmosphere containing 5% CO<sub>2</sub>. Cells were subcultured bi-weekly by trypsinization and the media was renewed every 2-3 days. MCF-7 cells were routinely cultured in IMEM (Improved Minimal Essential Media) containing phenol red, and 2 mM glutamine (Biofluids) supplemented with 10% FCS.

Hormone reagents 17 β-estradiol, progesterone, and *trans* 4'-hydroxytamoxifen were obtained from Sigma. The anti-estrogen ICI 162,780 was obtained from Alan Wakeling (ICI

Pharmaceuticals). Stock solutions of each steroid were prepared in absolute ethanol and diluted directly into media.

The hormone stimulation procedure was an adaptation of the procedure described elsewhere (May and Westley, 1986). Experimental media for BT-483 cells consisted of phenol red free RPMI 1640 (Life Technologies) supplemented with 2 mM glutamine, 20% CCS, 10µg/ml bovine insulin and 1X antibiotic/antimycotic solution. BT-483 cells were plated at a density of  $3 \times 10^6$  cells per T75 flask (Costar) in phenol red containing media. At 70-80% confluency, cells were depleted of steroids as previously described (May and Westley, 1986). Experimental media for MCF-7 cells was phenol red free IMEM (Biofluids) supplemented with 2 mM glutamine, 5% CCS and 1X antibiotic/antimycotic solution. Cells were plated at  $5 \times 10^6$  cells per 150 mm plate (Corning) in phenol red free IMEM for 5-6 days before the refeeding with fresh media containing steroids at defined concentrations. Fetal calf serum used in hormone studies (CCS) was stripped of endogenous estrogens with dextran coated charcoal as described elsewhere (May and Westley, 1986). Dextran T-70 was obtained from Pharmacia, and acid washed, neutralized activated charcoal from Sigma.

Cycloheximide was obtained from Sigma and diluted in water to a stock concentration of 50 mM. Cycloheximide was added to culture media at a concentration of 50µM for 1 hour prior to the addition of 10 nM estradiol or 0.01% ethanol. Trypan blue was obtained from Sigma and the exclusion assay performed according to the manufacturer's protocol.

Analysis of the estrogen and progesterone receptor content of the BT-483 cell line was performed in parallel with a reference T-47D breast cancer cell line by Dr. David Zava (Aeron Biotechnology, Inc., San Leandro, California).

## 2. RNA Extraction and Northern Blotting

RNA was extracted from cells with guanidinium isothiocyanate as described elsewhere (Chirgwin *et al.*, 1979). Cytoplasmic RNA was isolated from BT-483 monolayers by a combination of NP-40 lysis and mechanical disruption (Sambrook *et al.*, 1989) before the addition of lysates to guanidinium isothiocyanate. Total RNA from breast cancer cell lines was subjected to electrophoresis and blotted as described (Sambrook *et al.*, 1989).

Northern blots were hybridized separately with probes for BRCA1 and BRCA2 and 18S. Since total RNA was electrophoresed and transferred for these blots, the 18S RNA levels accurately reflect the amount of total RNA loaded per lane. The probe for BRCA1 was a 620 bp gel purified PCR<sup>TM</sup> product obtained with oligonucleotide primers 4L and 4R (5'-TACCCTATAAGCCAGAATCCA-3' and 5'-GGCAAACCTTGACACGAGCA-3'; SEQ ID NO:112 and SEQ ID NO:113, respectively) that amplified base pairs 4506-5126 of the published sequence (Miki *et al.*, 1994). The BRCA2 probe was obtained by PCR<sup>TM</sup> amplification of genomic DNA with oligonucleotide primers 5'-GGTACTAGTGAAATCACCAGT-3' and 5'-GTGAATGCGTGCTACATTCAT (forward; SEQ ID NO:114 and reverse; SEQ ID NO:115, respectively) spanning base pairs 4880-5979 in exon 11 of the Genbank sequence (Accession # U43746, Tavtigian *et al.*, 1996). The 18S and 36B4 probes were obtained from the American Type Culture Collection (ATCC #77242 and # 65917). Probes were labeled by random hexanucleotide extension (Feinberg and Vogelstein, 1983) with <sup>32</sup>P dCTP (Amersham).

Blots were hybridized at 42°C in 50% formamide solution containing dextran sulfate (Oncor) for 48 hours and subjected to a final wash in 0.5X SSC, 0.1% SDS at 65°C. Hybridization signals were quantitated by direct exposure to a PhosphorImager screen using Imagequant software supplied by the manufacturer (Molecular Dynamics). BRCA1 and BRCA2 were exposed to the PhosphorImager screen overnight and then exposed to x-ray film; 18S was exposed for 20 minutes to the PI screen and 2 hours to film.

#### B. Induction of BRCA1 and BRCA2 Expression in Breast Cancer Cells

Surges of the steroid hormones estrogen and progesterone occur during puberty (Drife, 1986), the menstrual cycle (Longacre and Bartow, 1986), and pregnancy (King, 1993). These surges profoundly change the proliferation, differentiation and architecture of the breast ductal epithelium from which the most common form of breast cancer arises (Slii *et al.*, 1994). Ductal carcinomas that are estrogen receptor positive depend on estrogen as an adjuvant to uncontrolled growth (King, 1993); however, these tumors are more differentiated, have a better prognosis (McGuire *et al.*, 1992) and are more likely to regress with antiestrogen therapy than are estrogen receptor negative tumors. Breast tumors from women less than 40 years old have a higher rate

of proliferation, are more aggressive and are more likely to be estrogen receptor negative than tumors from postmenopausal women (Marcus *et al.*, 1994).

Estrogen modulates growth and differentiation of human breast epithelium (Drife, 1986); however, the exact pathway by which it exerts its proliferative effects has not been elucidated. Estrogen combines with the estrogen receptor to modulate the transcription of a specific subset of genes that include autocrine and paracrine polypeptide growth factors such as IGF-1, TGF- $\alpha$ , and PDGF (Kasid and Lippman, 1987), the progesterone receptor (Horwitz and McGuire, 1978) and oncogenes such as c-myc (Dubik *et al.*, 1987). It has been previously demonstrated that steroid hormones regulate BRCA1 expression in human breast cancer cell lines (Spillman and Bowcock, 1995; Gudas *et al.*, 1995). *In vivo* data for murine BTCA1 also demonstrates that the highest levels of BRCA1 expression are observed in rapidly proliferating cells and in tissues that are sensitive to steroid hormones, such as the mammary gland (Marquis *et al.*, 1995 and Lane *et al.*, 1995).

The effect of steroid hormones on BRCA1 and BRCA2 mRNA expression was examined in the estrogen receptor positive breast cancer cell lines BT-483 and MCF-7. BT-483 cells were cultured in estrogen depleted phenol-red free media for 5 days before being switched to media containing 17  $\beta$ -estradiol and/or progesterone for an additional five days. The effect of estrogen or progesterone on BRCA1 and BRCA2 mRNA expression in BT-483 cells were performed in triplicate and BRCA1 and BRCA2 expression was quantified relative to the ethanol control.

Expression of both BRCA1 and BRCA2 mRNAs was suppressed in cells cultured in steroid depleted media. A striking elevation of BRCA1 and BRCA2 steady-state mRNA levels could be seen after five days of estrogen stimulation. In addition to the major BRCA1 transcript of 7.8 kb, an additional minor transcript of approximately 4 kb was also induced by estrogen in a similar fashion. Estrogen upregulated BRCA1 expression by approximately 17 fold and BRCA2 expression by approximately 50 fold. Similar results were seen in MCF-7 cells after severe serum deprivation.

A classic effect of estrogen on breast cancer cells is its ability to increase expression of the progesterone receptor (Horwitz and McGuire, 1978). In BT-483 cells estrogen acts *via* an active estrogen receptor to induce both progesterone receptor mRNAs and protein; however, progesterone alone failed to induce BRCA1 or BRCA2 mRNA expression in BT-483 and MCF-7 cells and the combination of estrogen and progesterone was neither synergistic nor completely antagonistic.

Both the BRCA1 and BRCA2 steady-state mRNA levels are both substantially elevated after estrogen treatment in the BT-483 and MCF-7 breast cancer cell lines. The finding that BRCA2 mRNA levels were also elevated by estrogen was initially surprising. BRCA2 mutations are thought to contribute to a significant proportion of male breast cancers (Wooster *et al.*, 1994) in addition to causing female breast cancers. Mutations in the androgen receptor have been shown to be responsible for some cases of male breast cancer (MacLean *et al.*, 1995), and the effect of the steroid hormone testosterone on the regulation of BRCA1 and BRCA2 mRNA levels is not known. However, estrogen may regulate BRCA1 and BRCA2 in male breast cancers as well, because male breast cancers are more likely to be estrogen receptor positive than female breast cancers (Hecht and Winchester, 1994). In terms of histology, female and male breast carcinomas are indistinguishable (Hecht and Winchester, 1994).

The BT-483 breast cancer cell line was derived from a 23 year old woman with breast cancer (Lasfargues *et al.*, 1978). BT-483 cells grow very slowly in culture. The doubling time of these cells is approximately 120 hours (Lasfargues *et al.*, 1978), which is similar to the time needed for tumor doubling *in vivo* (Rew *et al.*, 1992). These cells are exquisitely sensitive to estrogen, and will cease proliferation in a rich media containing steroid depleted serum (20% charcoal-stripped serum + insulin with a media change every day). In contrast, steroid deprivation of MCF-7 cells requires more drastic conditions (a very minimal media that is not changed for the first five days prior to the addition of estrogen). This treatment slows MCF-7 cell proliferation significantly and is required to demonstrate elevation of BRCA1 and BRCA2 mRNAs by estrogen.

Failure of progesterone to affect levels of BRCA1 or BRCA2 in response to estrogen in either BT-483 or MCF-7 cells is interesting because in normal breast development both estrogen



and progesterone are needed to complete the proliferation and differentiation of the breast tissue. Estrogen regulates the development of the ductules and progesterone regulates the development of the lobules (King, 1993). In women with germline BRCA1 mutations, although most tumors are of ductal origin, some are of lobular origin, mimicking the pattern seen in sporadic cases  
5 (Marcus *et al.*, 1994).

Studies on proliferation of breast cancer cell lines, with combinations of estrogen and progesterone do not give such clear results (King, 1993) although progestins predominately inhibit the estrogen-induced proliferation of breast cancer cell lines (Clarke and Sutherland,  
10 1990). In a previous study (Gudas *et al.*, 1995), progesterone was able to induce BRCA1 expression in the T-47D breast cancer cell line. However, this T-47D cell line is unusual because expression of the progesterone receptor is approximately 85 times greater in T-47D cells than in the BT-483 cells. Classic estrogen receptor positive breast cancer cell lines such as MCF-7 (Horwitz and McGuire, 1978) and BT-483 depend on estrogen induction of the  
15 progesterone receptor. Gudas *et al.* did not investigate the regulation of BRCA1 by progesterone in the MCF-7 cell line. Herein is evidence for the primary hormone controlling the elevation of BRCA1 and BRCA2 mRNAs being estrogen, not progesterone.

BRCA1 and BRCA2 are both tumor suppressor genes. Inactivation of these genes in  
20 women with germline mutations is frequently by deletions revealed by a loss of heterozygosity in the tumor (Merajver *et al.*, 1995 and Gudmundsson *et al.*, 1995). A few families with BRCA1 linked breast cancer do not have alterations in the coding sequences of BRCA1, raising the possibility of mutations in regions controlling BRCA1 expression. In the absence of coding mutations in the BRCA1 in sporadic breast tumors, alterations in the regulation of BRCA1  
25 expression are presumed to contribute to the cancerous phenotype (Thompson *et al.*, 1995). Failure to induce the postulated estrogen responsive protein or alterations in the regulatory pathway involving elevation of BRCA1 and BRCA2 mRNAs could result in a novel mechanism of malignant transformation through the loss of BRCA1 or BRCA2 transcripts.

### 30 C. Blocking of Estrogen Induction of BRCA1 and BRCA2 by Antiestrogens

Effects of estrogen mediated through the estrogen receptor can be competitively inhibited by antiestrogenic compounds. Two major classes of estrogen antagonists are

nonsteroidal antiestrogens such as *trans* 4'-hydroxytamoxifen (4-OHT) and steroidal antiestrogens such as ICI 182,780 (Wakeling *et al.*, 1989). While both classes of antiestrogens compete for binding to the estrogen receptor, they exert different actions on the activation of the estrogen receptor. Steroidal antiestrogens appear to prevent binding of the estrogen receptor to DNA while nonsteroidal antiestrogens fail to activate the estrogen-inducible transactivating function of the estrogen receptor protein (Green, 1990).

To confirm that the induction of BRCA1 and BRCA2 mRNA expression by estrogen was mediated by the estrogen receptor, the steroidal antiestrogen ICI 182,780 and the nonsteroidal antiestrogen *trans* 4'-hydroxytamoxifen were used in a competitive inhibition study. The results were analyzed by northern blotting. BT-483 cells were cultured as described previously with varying amounts of estrogen and antiestrogen. The antiestrogens ICI 182,780 and 4-OHT do not induce BRCA1 or BRCA2 expression. The expected estrogen mediated induction of BRCA1 and BRCA2 mRNAs is seen in the absence of any antiestrogen. When the amount of estrogen was held constant and the amount of antiestrogen varied, it was found that a one hundred fold molar excess of the antiestrogen ICI 182,780 was required to inhibit the estrogen induction of BRCA1 and BRCA2 mRNAs and to return their mRNA levels to a baseline level. Interestingly, a one hundred fold excess of ICI 182,780 is also the amount reported to be needed to block breast cancer cell proliferation *in vivo* in the presence of estradiol (Wakeling *et al.*, 1991).

Similar results were achieved with the *trans* 4'-hydroxytamoxifen. A one hundred fold excess of 4-OHT sharply reduced the amount of BRCA2 and BRCA1 mRNAs. The ability of two different classes of antiestrogen to block the expression of BRCA1 and BRCA2 in the presence of estradiol confirms that the expression of these genes is mediated by the estrogen receptor.

#### **D. Time Frame of Estrogen Induction of BRCA1 and BRCA2 mRNA**

The time at which BRCA1 and BRCA2 steady-state mRNA levels were elevated after estrogen stimulation was investigated in BT-483 cells. Cells were treated with estrogen and cytoplasmic RNA was isolated at varying times. Northern blot analysis of RNA obtained at regular time intervals for a total of ninety-six hours revealed that the initial expression levels of

BRCA1 were negligible and remained so during the first 18 hours following estrogen stimulation. The sharp elevation of BRCA1 mRNA between 18 and 24 hours after initial estrogen stimulation was particularly striking. This elevation persisted with continued estrogen stimulation and mRNA levels remained elevated for at least 96 hours.

5

The time and pattern of elevation of BRCA2 mRNA steady-state levels was remarkably similar to that demonstrated for BRCA1 mRNA. Levels of BRCA2 mRNA were negligible until 24 hours, at which time a sharp increase in the amount of BRCA2 transcript was detected. The increase in BRCA2 mRNA remained constant to 96 hours and was not subject to  
10 downregulation in the continued presence of estrogen. A continuous presence of estrogen was not necessary for the induction of BRCA1 and BRCA2 mRNA expression. A limited 9 hour pulse of estrogen chased with steroid depleted media was sufficient to induce BRCA1 and BRCA2 mRNA expression in cells harvested 24 hours after initiation of estrogen stimulation.

15

The response of BRCA1 and BRCA2 to estrogen occurs at the same time. In the BT-483 cell line mRNA levels of both genes are elevated 18 to 24 hours after estrogen stimulation, suggesting that they may have been coordinately regulated. This may be because they both play a role in control of the cell cycle. BRCA1 has been postulated to control cell proliferation and to maintain the cell in a differentiated state (Marcus *et al.*, 1994). Recent data (Vaughn *et al.*, 1996  
20 and Gudas *et al.*, 1996) indicate that the highest levels of BRCA1 mRNA and protein are seen in late G1 and early S phase, suggesting a role for BRCA1 in cell cycle regulation. Elevation of cyclin D1 mRNA has been observed at the same time as BRCA1 and BRCA2 mRNAs are elevated, supporting this hypothesis.

25

#### **E. Blocking of BRCA1 and BRCA2 Estrogen Induction by Cycloheximide**

The time lag between the initiation of estrogen stimulation and the increase in BRCA1 and BRCA2 mRNA levels suggests that estrogen acts indirectly on these two genes. If prior synthesis of intermediate proteins is necessary, then treatment of cells with the protein inhibitor cycloheximide should block the observed increase in BRCA1 and BRCA2 mRNA levels  
30 following estrogen treatment.

Cells were pretreated with cycloheximide for 1 hour prior to the addition of estrogen and harvested after 24 hours of estrogen stimulation. All studies were done in triplicate. Treatment with cycloheximide before the addition of estrogen completely blocked the increase in BRCA1 and BRCA2 mRNA levels. Cells treated with cycloheximide and no estrogen, as well as cells  
5 treated with no cycloheximide and no estrogen, did not result in an increase in BRCA1 and BRCA2 mRNA levels. Cells treated with estrogen and no cycloheximide demonstrated the expected previously observed increase in BRCA1 and BRCA2 mRNA levels. The effect of extended incubation with cycloheximide on cell viability was assayed by trypan blue exclusion and did not differ significantly between control and experimental cells, implying that the  
10 cycloheximide effect was not due to cell death. The ability of cycloheximide to block the induction of BRCA1 and BRCA2 was not due to a generalized decrease in transcription, because expression of the constitutively expressed estrogen-independent 36B4 mRNA (ribosomal phosphoprotein P0; Masiakowski *et al.*, 1982) showed no significant difference between cycloheximide treated and untreated cells.

15 The effect of estrogen on BRCA1 and BRCA2 steady-state mRNA levels by estrogen is indirect and requires prior protein synthesis as demonstrated by the action of cycloheximide. The implication of this is that an estrogen inducible protein may coordinately elevate the levels of BRCA1 and/or BRCA2 mRNAs. Alternatively, these genes may be induced by distinct  
20 estrogen induced pathways.

#### EXAMPLE XIV

##### BARD1 And BRCA1 In Discrete Nuclear Domains

25 The *BRCA1* tumor suppressor has been implicated in familial cases of early-onset breast and ovarian cancer (Hall *et al.*, 1990; Miki *et al.*, 1994). However, the biochemical functions of its protein product are not defined and the mechanism by which it counters tumor formation during normal development is not understood. The major isoform of BRCA1 is a polypeptide of ~220 kilodaltons that bears several recognizable amino acid motifs: these include a zinc-binding  
30 RING domain that lies near the amino terminus, two nuclear localization signals, and two tandem copies of the BRCT motif that reside at the carboxy-terminus (Miki *et al.*, 1994; Chen

*et al.*, 1996a; Thakur *et al.*, 1997; Koonin *et al.*, 1996). As described herein above, BRCA1 associates *in vivo* with BARD1. The interaction between these proteins is abolished by tumorigenic missense mutations in the RING domain of BRCA1, suggesting that tumor suppression may be mediated by a heteromeric complex of BRCA1 and BARD1.

5

Products of the *BRCA1* gene are found in a broad spectrum of cell and tissue types (Miki *et al.*, 1994; Lane *et al.*, 1995; Marquis *et al.*, 1995); however, its expression in most (Chen *et al.*, 1996c; Vaughn *et al.*, 1996a, Gudas *et al.*, 1996; Rajan *et al.*, 1996), but not all (Aprelikova *et al.*, 1996), cell types is tightly regulated during cell cycle progression. In resting  
10 cells, the levels of BRCA1 transcripts and polypeptides are either low or undetectable. However, after these cells receive a mitotic stimulus the steady-state levels of BRCA1 products rise in late G1, peak just prior to the onset of DNA synthesis, and persist for the duration of S phase and most of M phase. In addition, BRCA1 polypeptides become hyperphosphorylated as they begin to accumulate in late G1 (Chen *et al.*, 1996c). While not conclusive, these findings  
15 suggest that BRCA1 may be involved in some aspect of cell cycle regulation (Chen *et al.*, 1996c; Vaughn *et al.*, 1996a; Gudas *et al.*, 1996; Rajan *et al.*, 1996).

Recent studies indicate that BRCA1 resides predominately in the nuclei of normal cells (Chen *et al.*, 1995; Scully *et al.*, 1996; Chen *et al.*, 1996b; Thomas *et al.*, 1996). During S  
20 phase, when their levels are most abundant, BRCA1 polypeptides exist in distinct subnuclear bodies, termed BRCA1 nuclear dots. Although the function of these dots is not known, most, but not all, co-stain with antibodies that recognize HsRad51, a DNA-binding protein that shares extensive homology with the yeast Rad51 and *E. coli* RecA proteins (Scully *et al.*, 1997). HsRad51 promotes homologous pairing and single strand exchange between DNA duplexes, and  
25 it has been implicated in a variety of nuclear processes, including DNA recombination, RNA transcription and DNA repair (Scully *et al.*, 1997 for additional references). As such, the co-localization of BRCA1 and HsRad51 to the same subnuclear structures provides important clues about BRCA1 function (Scully *et al.*, 1997).

30 To obtain additional insights into the function of BRCA1, the expression and subcellular distribution of BARD1 was examined during cell cycle progression. In contrast to BRCA1, the steady-state levels of BARD1 remain relatively constant throughout the cell cycle. Subcellular

fractionation of synchronized cell populations showed that BARD1 resides in the nuclei of proliferating cells, and two-color immunofluorescence with BARD1-specific antibodies revealed a punctate pattern of nuclear staining with nearly perfect co-localization of BARD1 and BRCA1. However, the punctate pattern of BARD1 immunostaining was observed in S-phase, but not in G1-phase, cells. Therefore, despite the presence of BARD1 polypeptides in the nucleus throughout cell cycle progression, their accumulation into BRCA1 nuclear dots is an S phase-specific phenomenon that may require recruitment by BRCA1. This cell cycle-dependent co-localization of BARD1 and BRCA1 further indicates a role for BARD1 in BRCA1-mediated tumor suppression.

### 1. Experimental Materials

HBL-100 and T24 cell lines were obtained from the American Type Culture Collection and normal human mammary epithelial cells (HMECs) were purchased from Clonetics Corp. (San Diego, CA). Three different BARD1-specific antibody reagents were used in this study: a mouse polyclonal antiserum, a mouse monoclonal antibody, and an affinity-purified rabbit polyclonal antiserum. To prepare the latter, a cDNA fragment of human BARD1 was inserted into the *Bam*HI/*Hind*III sites of the pMAL-c2 bacterial expression vector (New England Biolabs, Beverly, MA); the resultant plasmid encodes MBP-EE, a hybrid polypeptide comprised of the *E. coli* maltose binding protein (MBP) fused to residues 141-388 of BARD1. MBP-EE polypeptides were then purified from *E. coli* lysates by affinity chromatography on an amylose resin (New England Biolabs) and conjugated to CNBr-activated Sepharose 4B (Pharmacia Biotech). The rabbit polyclonal antiserum raised against GST-EE, a hybrid polypeptide containing silkworm GST fused to residues 141-388 of BARD1, was then purified by sequential affinity chromatography on HiTrap protein A-Sepharose (Pharmacia Biotech) and MBP-EE-conjugated Sepharose 4B. The BARD1-specific mouse polyclonal antiserum and monoclonal antibody were raised by immunizing mice with the GST-EE polypeptide. The monoclonal antibody was used for BARD1 immunoblots (e.g., FIGs. 1 and 5). Monoclonal antibodies that recognize BRCA1 (MS110), cyclin A (Ab-3), NuMA (Ab-1), and  $\alpha$ -tubulin (Ab-1) were purchased from Oncogene Research Products. The CDK2-specific antiserum (M2) was obtained from Santa Cruz Biotechnology.

## 2. Steady-State Levels of BARD1 Remain Constant During Cell-Cycle Progression

To compare the expression of BARD1 and BRCA1 polypeptides with respect to the cell cycle, their steady-state levels were measured in synchronized populations of T24 bladder carcinoma cells. T24 cells were arrested in G0 by contact inhibition in 175 cm<sup>2</sup> flasks. After at least 3 days of confluence, the cells were split 1:10 by seeding multiple 100 mm dishes at a concentration of  $\sim 10^6$  cells/dish. Individual cultures were harvested at various times after replating (Chen *et al.*, 1996c). The cell cycle distribution profile of each culture was then determined by FACS analysis and protein levels were evaluated by immunoblotting.

Ten dishes were harvested at each timepoint after replating - two for FACS analysis and eight for Western analyses. To determine the cell cycle distribution at each timepoint, the contents of each dish were incubated for 10 min at room temperature in 2 ml of trypsin/EDTA solution (0.25% trypsin, 0.1% EDTA in HBSS w/o CaMg; Mediatech, Inc.). The trypsinized cells were then washed in 10 ml of growth medium (McCoy's 5A, 10% FBS) and resuspended in 1.5 ml of ice-cold PBS (w/o CaMg). After adding 3.5 ml of ice-cold 100% ethanol dropwise, the cells were fixed at 4°C for at least 16 h. The fixed cells were pelleted, resuspended in 1 ml of PI staining solution (50 µg/ml propidium iodide, 100 U/ml RNase A, 0.1% glucose in PBS w/o CaMg), incubated for at least 1h at room temperature, and analyzed on a FACScan flow cytometer (Becton Dickinson).

For Western analyses the contents of eight dishes were lysed in a total of 300 µl RIPA buffer (50 mM Tris pH 7.6, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS) containing complete protease inhibitor cocktail (Boehringer Mannheim) and phosphatase inhibitors (5 mM β-glycerophosphate, 10 mM benzamidine, and 0.5 mM sodium orthovanadate). The lysate was vortexed for 10 min at 4°C and cleared of insoluble debris by centrifugation for 10 min at 12,000 RPM in a microfuge at 4°C. The protein concentration of each supernatant was determined using the BCA Protein Assay Reagent (Pierce). Equivalent aliquots of each lysate were subjected to Western analyses with antibodies specific for CDK2, cyclin A, BRCA1, and BARD1. Western analyses were conducted by enhanced chemiluminescence (Amersham) using 80 µg of lysate for BRCA1 immunoblots and 30 µg for CDK2, cyclin A, and BARD1 immunoblots.

The cells display the expected expression patterns for known cell cycle regulatory molecules. For example, CDK2 is present throughout the cell cycle and its steady state levels increase modestly in S and G2/M cells. However, the levels of its regulatory subunit, cyclin A, rise dramatically after the G1/S transition. BRCA1 shows an expression profile similar to that described in a previous study of T24 cells (Chen *et al.*, 1996c); thus, while few, if any, BRCA1 products are detected in resting or G1 cells, BRCA1 expression increases markedly as cells enter S phase. In contrast, comparable levels of BARD1 polypeptides are seen at all timepoints, indicating that BARD1 expression remains relatively constant throughout the cell cycle. In addition, Western analysis of subcellular fractions from synchronized cell populations demonstrate that BARD1 remains in the nuclear compartment of G1- and S-phase proliferating cells.

### 3. BARD1 Polypeptides Reside In Discrete Subnuclear Bodies

The subcellular distribution of BARD1 polypeptides was evaluated by immunofluorescent staining of unsynchronized HBL-100 cells, a human line of normal mammary epithelial cells that was presumably immortalized by transforming sequences of the SV40 papovavirus (Caron de Fromentel *et al.*, 1985). A mouse polyclonal antiserum was prepared against residues 141-388 of BARD1, a segment that bears no homology to other known proteins. Approximately  $2.5 \times 10^6$  cells were seeded onto microscope slides in a 150 mm culture dish. After 2 days, the cells were fixed with 4% paraformaldehyde for 15 min and permeabilized in 0.2% Triton X-100 for 10 min. Non-specific staining was blocked by a 60 min incubation with 2% bovine serum albumin in phosphate-buffered saline (BSA/PBS solution) and two 15 min treatments with the Avidin/Biotin Blocking Kit (Vector Laboratories, Burlington, CA). After a 60 min incubation with primary antibody, the cells were treated with 8  $\mu\text{g/ml}$  biotinylated secondary antibody (Vector Laboratories) for 45 min and 20  $\mu\text{g/ml}$  fluorescein avidin D (Vector Laboratories) for an additional 30 min. The cells were then treated with 100  $\mu\text{g/ml}$  of RNase A in PBS for 20 min at 37°C, and with 10  $\mu\text{g/ml}$  propidium iodide in PBS for an additional 20 min. The stained cells were mounted under coverslips with VECTASHIELD mounting medium (Vector Laboratories) and sealed with nail polish. Immunofluorescence was recorded using a confocal microscope equipped with a MRC-1024 Lasersharp confocal imaging system (Bio-Rad Laboratories). All the above procedures were performed at room temperature except where indicated.



After staining with either the BARD1-specific antiserum or a BRCA1-specific monoclonal antibody (MS110; Oncogene Research Products; Scully *et al.*, 1996), the cells were counter-stained with propidium iodide to highlight the nuclei. A characteristic pattern of BRCA1 subcellular distribution was observed in which BRCA1 nuclear dots appeared in some, but not all, interphase cells. Likewise, the BARD1-specific antiserum generated a similar pattern of punctate nuclear staining in a subset of interphase cells. The same results were also obtained using T24 colon carcinoma cells and primary human mammary epithelial cells.

#### 4. BARD1-Containing Nuclear Foci Appear Specifically In S-Phase Cells

The nuclear dot pattern of BRCA1 staining has been shown to arise specifically during S-phase of the cell cycle (Scully *et al.*, 1997). To determine whether the subnuclear structures that stain with BARD1 have a similar cell-cycle dependence, synchronized populations of T24 cells were stained with BARD1- or BRCA1-specific monoclonal antibodies. Cells harvested at 8 h (91% G1 phase cells) and 20 h (56% S phase cells) after replating were analyzed. In some studies the monoclonal antibodies were pre-absorbed with an excess of the BRCA1 immunogen (GST-BR304), the BARD1 immunogen (GST-EE) or the parental GST polypeptide.

Cells bearing BRCA1 nuclear dots were abundant in the S phase population but were rarely observed in the G1 population. The specificity of BRCA1 staining was confirmed in blocking studies in which the primary antibody was preabsorbed with a resin-bound protein containing silkworm-glutathione S-transferase (GST) fused to the amino-terminal 304 residues of BRCA1 - the same BRCA1 moiety used to generate the MS110 monoclonal antibody (Scully *et al.*, 1996). As expected, staining of BRCA1 nuclear dots was completely abolished by preabsorption with the GST-BRCA1 fusion protein but not with the parental GST polypeptide.

Immunofluorescence analysis of synchronized cell populations revealed that the appearance of BARD1-staining foci with respect to the cell cycle resembles that of BRCA1 nuclear dots. Thus, these foci are present in most cells of the S phase population (panel f) but not the G1 population. Moreover, the staining of S phase cells with BARD1-specific antibodies was ablated by preabsorption with GST-EE, a polypeptide containing GST fused to residues

141-388 of BARD1, but not by GST itself. These data show that the BARD1-staining nuclear structure arises in an S phase-specific fashion reminiscent of the BRCA1 nuclear dots.

#### 5. BARD1 And BRCA1 Polypeptides Co-Localize In BRCA1 Nuclear Dots

5 If BARD1 is a physiologically-relevant partner of BRCA1 then the two proteins should reside in the same subcellular structures. Therefore, to determine whether the S-phase nuclear foci recognized by BARD1- and BRCA1-specific antibodies are one and the same, two-color immunofluorescence studies were conducted by staining HBL-100 cells simultaneously with an affinity-purified BARD1-specific rabbit antiserum and a mouse monoclonal antibody that  
10 recognizes either BRCA1 or PML; the latter is a RING protein that resides in distinct subnuclear structures referred to as PML oncogenic domains (PODs) (Dyck, *et al.*, 1994; Koken *et al.*, 1994).

Cells were incubated simultaneously with the two primary antibodies for 60 min. After  
15 treatment with Texas Red-conjugated anti-rabbit goat IgG (Vector Laboratories) and biotinylated anti-mouse goat IgG (Vector Laboratories) for 45 min, the cells were incubated for an additional 30 min with fluorescein avidin D. The immunostained cells were then mounted as described above (without RNase A digestion and propidium iodide staining). A 10 µg aliquot of the BARD1- or BRCA1-specific monoclonal antibody was preabsorbed by overnight incubation  
20 at 4°C with 50 µg of either the parental GST polypeptide or the cognate immunogen (GST-EE or GST-BR304, respectively) immobilized on glutathione-agarose beads. Images of BARD1-staining (red) and BRCA1- or PML-staining (green) from the same cells were then collected both separately and conjointly.

25 BARD1-staining coincides almost perfectly with the BRCA1 nuclear dots of HBL-100 cells. In contrast, BARD1-staining structures are distributed randomly with respect to the PML-oncogenic domains. Similar results were obtained in two-color immunofluorescence studies of normal human mammary epithelial cells. These data demonstrate that BARD1 specifically co-localizes with BRCA1 in the same subnuclear bodies. Co-localization of BRCA1 and  
30 BARD1 in nuclear dots appears to be independent of cell type and the degree of neoplastic transformation.

## 6. Subcellular Distribution Of BARD1 Polypeptides During Cell Cycle Progression

The BARD1-staining nuclear foci are only apparent by immunofluorescence microscopy after the onset of S phase. Nevertheless, Western analysis of lysates from synchronized cell populations show that the steady-state levels of BARD1 polypeptides remain relatively constant throughout the cell cycle. To address the question of where BARD1 resides in G1-phase cells, the subcellular distribution of BARD1 polypeptides was initially examined by Western analyses of nuclear, cytoplasmic, and membrane fractions prepared from asynchronous T24 cells.

To prepare whole cell lysates of unsynchronized T24 cells, the cellular contents of two 150 mm dishes ( $\sim 1.7 \times 10^7$  cells/dish) were lysed in 1 ml of RIPA buffer (containing protease and phosphatase inhibitors, as described above). Whole cell lysates of synchronized T24 cells (8h or 20 h after replating) were prepared by lysing the contents of six 150 mm dishes in 1 ml of RIPA buffer ( $\sim 2.6 \times 10^6$  cells/dish). Each whole cell lysate was vortexed for 15 min at 4°C and cleared of insoluble debris by centrifugation for 10 min at 12,000 RPM in a microfuge at 4°C. To prepare membrane, cytoplasmic, and nuclear fractions from unsynchronized cells, the contents of seven 150 mm dishes ( $\sim 1.7 \times 10^7$  cells/dish) were resuspended in 5 ml of hypotonic lysis buffer and processed as described (Abrams *et al.*, 1982).

For synchronized cells, the contents of twenty-five 150 mm dishes ( $\sim 2.6 \times 10^6$  cells/dish) were resuspended in 5 ml of hypotonic lysis buffer and processed to prepare subcellular fractions (Abrams *et al.*, 1982). For detection of BRCA1, equivalent volumes of each fraction (corresponding to 300 µg of whole cell lysate) were immunoprecipitated with the BRCA1-specific rabbit antiserum and the immunoprecipitates were subjected to Western analysis with the BRCA1-specific MS110 monoclonal antibody. For detection of BARD1, NuMA, and  $\alpha$ -tubulin, equivalent volumes of each fraction (corresponding to 10 µg of whole cell lysate) were directly evaluated by Western analysis with the appropriate monoclonal antibody.

BARD1 and BRCA1 were concentrated in the nuclear fraction along with the nuclear matrix protein NuMA (Lydersen *et al.*, 1980). In contrast,  $\alpha$ -tubulin was found exclusively in the cytoplasmic and membrane compartments, indicating that there was little, if any, cross-contamination of the nuclear compartment with cytosolic proteins. Identical results were obtained by Western analysis of subcellular fractions from synchronized populations of T24

cells harvested at 8 h (98% G1 cells) and 20 h (52% S phase cells) after release from cell cycle arrest. Hence, during G1 phase of the cell cycle, when BARD1 is not found in BRCA1 nuclear dots by immunofluorescent staining, the analysis of subcellular fractions reveals it to be predominantly a nuclear protein.

5

Immunostaining with BARD1-specific antibodies was not observed in G1 cells, despite the fact that BARD1 polypeptides were readily detected by Western analysis of nuclear fractions derived from these cells. Several explanations can be invoked to account for this phenomenon. For example, the epitopes recognized by the BARD1-specific antibodies may be masked during certain stages of the cell cycle by interactions with other macromolecules. However, identical results were obtained with three different reagents raised against a substantial segment of human BARD1 (residues 141-388): an affinity-purified rabbit polyclonal antiserum, a mouse polyclonal antiserum, and a mouse monoclonal antibody. Furthermore, attempts to unmask hidden epitopes with heat or high salt did not elicit BARD1-specific staining in G1 cells, despite the fact that the monoclonal antibody readily detects denatured BARD1 polypeptides. Thus, a more plausible explanation for this phenomenon is that BARD1 polypeptides are distributed diffusely within the nuclei of G1 cells at concentrations too low for immunodetection. In contrast, the S phase-dependent accumulation of BARD1 into BRCA1 dots presumably increases their local concentration to levels detectable by immunofluorescence microscopy.

20

If BARD1 polypeptides are diffusely distributed in the nuclei of G1 cells, then all, or at least a significant subset, of these polypeptides must be recruited into the BRCA1 nuclear dots as cells progress into S phase. The re-localization of BARD1 may occur independently of BRCA1, or the BARD1 accumulation into the dots may require the prior formation of BRCA1/BARD1 heterodimers. In this regard, determination of the nuclear distribution of BARD1 in cells that lack functional BRCA1 will be feasible once cell lines are established from either *Brcal*-null mice or breast carcinomas of BRCA1 mutation carriers.

25

Germline mutations of either BRCA1 or BRCA2 are responsible for most cases of familial breast cancer. Thus, it is intriguing to note that these genes share a number of other similarities. First, unlike most tumor suppressors, BRCA1 and BRCA2 are rarely mutated in truly sporadic cases of breast cancer (Futreal *et al.*, 1994; Lancaster *et al.*, 1996; Teng *et al.*,

30

1996; Miki *et al.*, 1996). Second, the phylogenetic conservation of both genes is remarkably poor - for example, the mouse and human orthologs of their protein products exhibit only 58% identity at the amino acid level (Lane *et al.*, 1995; Abel *et al.*, 1995; Sharan *et al.*, 1995; Bennett *et al.*, 1995; Connor *et al.*, 1997; Sharan *et al.*, 1997). Third, the transcription of both genes is coordinately induced by estrogen (Example XIII, above). Fourth, the expression patterns of BRCA1 and BRCA2 with respect to the cell cycle are almost indistinguishable: both are induced in late G1 upon mitogenic stimulation of quiescent cells, and the levels of their gene products peak just prior to DNA synthesis (Chen *et al.*, 1996c; Vaughn *et al.*, 1996a; Gudas *et al.*, 1996; Rajan *et al.*, 1996; Vaughn *et al.*, 1996b; Wang *et al.*, 1997).

These intriguing parallels were underscored recently by the discovery that BRCA2 also interacts *in vivo* with HsRad51 (Sharan *et al.*, 1997; Mizuta *et al.*, 1997). Although the subcellular localization of BRCA2 has not yet been described, these findings suggest that BRCA1 and BRCA2 normally serve as components of a common biochemical pathway involving the HsRad51 protein (Scully *et al.*, 1997; Sharan *et al.*, 1997; Mizuta *et al.*, 1997). As such, the disruption of this pathway by mutations in BRCA1 or BRCA2 may be a critical step in the development of hereditary breast cancer. The specific localization of BARD1 into the BRCA1 nuclear dots of S phase cells suggests that it too may be an essential component of a HsRad51-associated pathway of tumor suppression.

\* ..... \*

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to

those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

## REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

5

Abbondanzo *et al.*, *Breast Cancer Res. Treat.*, 16:182(#151), 1990.

Abel, Xy, Yin, Lyons, Meisler, Weber, *Hum. Mol. Genet.*, 4:2265-2273, 1995.

Abrams, Rohrschneider, Eisenman, *Cell*, 29:427-439, 1982.

Allred *et al.*, *Breast Cancer Res. Treat.*, 16:182(#149), 1990.

10 Altschul, Gish, Miller, Myers, Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, 215:403-410, 1990.

Andersson, Davis, Dahlback, Jornvall, Russell, "Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme," *J. Biol. Chem.*, 264:8222-8229, 1989.

15 *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988.

Aprelikova, Kuthiala, Bessho, Etheir, Liu, *Oncogene*, 13:2487-2491, 1996.

Beaudet and Tsui, "A suggested nomenclature for designating mutations," *Hum. Mut.*, 2:245-248, 1993.

Bennett *et al.*, "Isolation of the mouse homologue of BRCA1 and genetic mapping to mouse  
20 chromosome 11," *Genomics*, 29:576-581, 1995.

Bork, "Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally," *Proteins: Structure, Function, and Genetics*, 17:363-374, 1993.

Bork *et al.*, "A superfamily of conserved domains in DNA damage-responsive cell cycle  
25 checkpoint proteins," *FASEB J.* 11:68-76, 1997.

Brown *et al.*, *Breast Cancer Res. Treat.*, 16:192(#191), 1990.

Brutlag *et al.*, *CABIOS*, 6:237-245, 1990.

Callebaut and Mornon, "From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair," *FEBS Letters* 400:25-30, 1997.

- Campbell, In: *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 13, Burden and Von Knippenberg, Eds. pp. 75-83, Amsterdam, Elsevier, 1984.
- Caron de Fromentel, Nardeux, Soussi, Lavialle, Estrade, Carloni, Chandrasekaran, Cassingena,  
5 *Exp. Cell Res.*, 160:83-94, 1985.
- Castilla *et al.*, "Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer," *Nature Genet.*, 8:387-391, 1994.
- Chapman and Verma, "Transcriptional activation by BRCA1," *Nature* 382:678-679, 1996.
- Chen, Chen, Riley, Allred, Chen, Hoff, Osborne, Lee, *Science*, 270:789-791, 1995.
- 10 Chen, Chen, Riley, Lee, Allred, Osborne, *Science*, 272:125-126, 1996b.
- Chen, Farmer, Chen, Jones, Chen, Lee, *Cancer Res.*, 56:3168-3172, 1996c.
- Chen, Li, Chen, Chen, Sharp, Lee, *J. Biol. Chem.*, 271:32863-32868, 1996a.
- Chien, Bartel, Sternglanz, Fields, "The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest," *Proc. Natl. Acad. Sci. USA*, 88:9578-  
15 9582, 1991.
- Chirgwin, Przybyla, MacDonald, Rutter, *Biochem.*, 18:5294-5299, 1979.
- Chou and Fasman, "Conformational Parameters for Amino Acids in Helical,  $\beta$ -Sheet, and Random Coil Regions Calculated from Proteins," *Biochemistry*, 13(2):211-222, 1974b.
- Chou and Fasman, "Empirical Predictions of Protein Conformation," *Ann. Rev. Biochem.*,  
20 47:251-276, 1978b.
- Chou and Fasman, "Prediction of b-Turns," *Biophys. J.*, 26:367-384, 1979.
- Chou and Fasman, "Prediction of Protein Conformation," *Biochemistry*, 13(2):222-245, 1974a.
- Chou and Fasman, "Prediction of the Secondary Structure of Proteins from Their Amino Acid Sequence," *Adv. Enzymol. Relat. Areas Mol. Biol.*, 47:45-148, 1978a.
- 25 Clarke and Sutherland, *Endocrin. Rev.*, 11:266-302, 1990.
- Clines *et al.*, "The structure of the human multiple exostoses 2 gene (EXT2) and characterization of homologs in mouse and *C. elegans*." *Am. J. Hum. Genet.*, In Press, 1997.
- Connor, Smith, Wooster, Stratton, Dixon, Campbell, Tait, Freeman, Ashworth, *Hum. Mol. Genet.*, 6:291-300, 1997.
- 30 Cropp *et al.*, "Loss of heterozygosity on chromosomes 17 and 18 in breast carcinoma: two additional regions identified." *Proc. Natl. Acad. Sci., U.S.A.*, 87:7737-7741, 1990.



- Dang *et al.*, "Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization," *Mol. Cell. Biol.*, 11:954-962, 1991.
- Drife, *Ann. New York Acad. Sci.*, 464:58-65, 1986.
- Dubik, Dembinski, Shiu, *Cancer Res.*, 47:6517-6521, 1987.
- 5 Durfee *et al.*, "The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit," *Genes Dev.*, 7:555-569, 1993.
- Dyck, Maul, Wilson, Miller, Chen, Kakizuka, Evans, *Cell*, 76:333-343, 1994.
- Easton, Bishop, Ford, Crockford, "Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium," *Am. J. Hum.*
- 10 *Genet.*, 52:678-701, 1993.
- Feinberg and Vogelstein, *Anal. Biochem.*, 132:6-13, 1983.
- Fetrow and Bryant, "New Programs for Protein Tertiary Structure Prediction," *Biotech.*, 11:479-483, 1993.
- Fields and Song, "A novel genetic system to detect protein-protein interactions," *Nature*,
- 15 340:245-246, 1989.
- Ford, Easton, Bishop, Narod, Goldgar, "Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium," *Lancet*, 343:692-695, 1994.
- Friedman *et al.*, "Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families," *Nature Genet.*, 8:399-404, 1994.
- 20 Futreal *et al.*, "*BRCA1* mutations in primary breast and ovarian carcinomas," *Science*, 266:120-122, 1994.
- Gefter *et al.*, *Somatic Cell Genet.* 3:231-236, 1977.
- Gemmill *et al.*, In: *Large insert cloning and analysis*, Dracopoli *et al.*, eds., John Wiley and Sons, Inc.
- 25 Goding, In: *Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, pp. 60-61, 65-66, 71-74, 1986.
- Green, *J. Steroid Biochem. Molec. Biol.*, 37:747-751, 1990.
- Guan and Dixon, "Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione *S*-transferase,"
- 30 *Analytical Biochemistry*, 192:262-267, 1991.
- Gudas, Li, Nguyen, Jensen, Rauscher III, Cowan, *Cell Growth and Diff.*, 7:717-723, 1996.
- Gudas, Nguyen, Li, Cowan, *Cancer Res.*, 55:4561-4565, 1995.

- Gudmundsson, Johannesdottir, Bergthorsson, Arason, Ingvarsson, Egilsson, Barkardottir, *Cancer Res.*, 55:4830-4832, 1995.
- Haber and Harlow, "Tumour-suppressor genes: evolving definitions in the genomic age," *Nature Genet.* 16:320-322, 1997.
- 5 Hall *et al.*, "Linkage of early-onset familial breast cancer to chromosome 17q21," *Science*, 250:1684-1689, 1990.
- Harper, Adami, Wei, Keyomarsi, Elledge, "The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases," *Cell*, 75:805-816, 1993.
- Hecht and Winchester, *Am. J. Clin. Pathol.*, 102(4 Suppl 1):S25-S30, 1994.
- 10 Hopp *et al.*, "A short polypeptide marker sequence useful for recombinant protein identification and purification," *Bio/Technology*, 6:1204-1210, 1988.
- Hopp, U.S. Patent 4,554,101.
- Horwitz and McGuire, *J. Biol. Chem.*, 253:2223-2228, 1978.
- Hosking *et al.* "A somatic BRCA1 mutation in an ovarian tumour (letter)". *Nature Genet.*, 15 9:343-344, 1995.
- Hsu, Wadman, Baer, "Formation of *in vivo* complexes between the TAL1 and E2A polypeptides of leukemic T cells," *Proc. Natl. Acad. Sci. USA*, 91:3181-3185, 1994.
- Jameson and Wolf, "The Antigenic Index: A Novel Algorithm for Predicting Antigenic Determinants," *Comput. Appl. Biosci.*, 4(1):181-186, 1988.
- 20 Johnson *et al.*, "Peptide Turn Mimetics," In: *Biotechnology and Pharmacy*, Pezzuto *et al.*, eds., Chapman and Hall, New York, 1993.
- Kasid and Lippman, *J. Steroid Biochem.*, 27:465-470, 1987.
- King, *Breast Cancer Res. Treat.*, 27:3-15, 1993.
- Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.
- 25 Kohler and Milstein, *Nature*, 256:495-497, 1975.
- Koken, Puvion-Dutilleul, Guillemin, Viron, Linares-Cruz, Stuurman, Jong, Szostecki, Calvo, Chomienne, Degos, Puvion, The, *EMBO J.*, 13:1073-1083, 1994.
- Koonin, Altschul, Bork, "BRCA1 protein products: functional motifs," *Nature Genet.*, 13:266-267, 1996.
- 30 Kyte and Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, 157(1):105-132, 1982.

- Lancaster *et al.* "BRCA2 mutations in primary breast and ovarian cancers." *Nature Genet.*, 13:238-240, 1996.
- Landschulz, Johnson, McKnight, "The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins," *Science*, 240:1759-1764, 1988.
- 5 Lane *et al.*, "Expression of *Brca1* is associated with terminal differentiation of ectodermally and mesodermally derived tissues in mice," *Genes Dev.*, 9:2712-2722, 1995.
- Langdon *et al.*, "Characterization and properties of nine human ovarian adenocarcinoma cell lines." *Cancer Res.*, 48:6166-6172, 1988.
- Lasfargues *et al.* "Isolation of two human tumor epithelial cell lines from solid breast carcinomas." *J. Natl. Cancer Inst.* 61:967-973, 1978.
- 10 Li *et al.*, "PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast and prostate cancer," *Science* 275:1943-1947, 1997.
- Liaw *et al.*, "Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome," *Nature Genet.* 16:64-67, 1997.
- 15 Lindblom *et al.*, "Loss of heterozygosity in familial breast carcinomas." *Cancer Res.*, 53:4356-4361, 1993.
- Longacre and Bartow, *Am. J. Surg. Pathol.*, 10:382-393, 1986.
- Lydersen and Pettijohn, *Cell*, 22:489-499, 1980.
- MacLean, Warne, Zajac, *Mol. Cell. Endo.*, 112:133-141, 1995.
- 20 Marcus, Watson, Page, Lynch, *J. Natl. Cancer Inst. Monographs Series*, 16:23-34, 1994.
- Marquis, Rajan, Wynshaw-Boris, Xu, Yin, Abel, Weber, Chodosh, *Nat. Genet.*, 11:17-26, 1995.
- Masiakowski, Breathnach, Bloch, Gannon, Krust, Chambon, *Nucleic Acids Res.*, 10:7895-7903, 1982.
- May and Westley, *Cancer Res.*, 46:6034-6040, 1986.
- 25 Mazoyer *et al.*, "A polymorphic stop codon in BRCA2," *Nature Genet.* 14:253-254, 1996.
- McGuire, Chamness, Fuqua, *J. Steroid Biochem. Molec. Biol.*, 43:243-247, 1992.
- Merajver *et al.*, "Somatic mutations in the BRCA1 gene in sporadic ovarian tumours," *Nature Genet.*, 9:439-443, 1995.
- Merajver, Frank, Xu, Pham, Calzone, Bennett-Baker, Chamberlain, Boyd, Garber, Collins, Weber, *Clin. Can. Res.*, 1:539-544, 1995.
- 30 Miki *et al.*, "Mutation analysis in the BRCA2 gene in primary breast cancers." *Nature Genet.*, 13:245-247, 1996.

- Miki *et al.*, "A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*," *Science*, 266:66-71, 1994.
- Miller, *Curr. Top. Microbiol. Immunol.*, 158:1, 1992.
- Mizuta, LaSalle, Cheng, Shinohara, Ogawa, Copeland, Jenkins, Lalande, Alt, *Proc. Natl. Acad. Sci. USA*, 94:6927-6932, 1997.
- 5 Monteiro *et al.*, "Evidence for a transcriptional activation function of *BRCA1* C-terminal region," *Proc. Natl. Acad. Sci. USA*, 93:13595-13599, 1996.
- Murre, McCaw, Baltimore, "A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins," *Cell*, 56:777-783, 1989.
- 10 Nakamura *et al.*, In: *Enzyme Immunoassays: Heterogeneous and Homogeneous Systems*, Chapter 27.
- Orita *et al.*, "Detection of polymorphisms of human DNA by gel electrophoresis as single-stranded conformation polymorphisms," *Proc. Natl. Acad. Sci., U.S.A.*, 86:2766-2770, 1989b.
- 15 Orita *et al.*, "Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction," *Genomics*, 5:874-879, 1989a.
- Phelan *et al.*, "Ovarian cancer risk in *BRCA1* carriers is modified by the *HRAS1* variable number of tandem repeat (VNTR) locus," *Nature Genet.*, 12:309-311, 1996.
- Rajan, Wang, Marquis, Chodosh, *Proc. Natl. Acad. Sci. USA*, 93:13078-13083, 1996.
- 20 Remington's *Pharmaceutical Sciences*, 15th Edition, pages 1035-1038 and 1570-1580.
- Rew, Campbell, Taylor, Wilson, *Br. J. Surg.*, 79:335-339, 1992.
- Sadowski, Bell, Broad, Hollis, "GAL4 fusion vectors for expression in yeast or mammalian cells," *Gene*, 118:137-141, 1992.
- Sambrook, Fritsch, Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989.
- 25 Saurin, Borden, Boddy, Freemont, "Does this have a familiar RING?" *Trends Biochem. Sci.*, 21:208-214, 1996.
- Scully, Chen, Plug, Xiao, Weaver, Feunteun, Ashley, Livingston, *Cell*, 88:265-275, 1997.
- Scully, Ganesan, Brown, Caprio, Cannistra, Feunteun, Schnitt, Livingston, *Science*, 272:123-125, 1996.
- 30 Sharan, Wims, Bradley, "Murine *Brcal*: sequence and significance for human missense mutations," *Hum. Molec. Genet.*, 4:2275-2278, 1995.

- Sharan, Morimatsu, Albrecht, Lim, Regel, Dinh, Sands, Eichele, Hasty, Bradley, *Nature*, (London), 1997.
- Shattuck-Eidens, "A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene: Implications for presymptomatic testing and screening," *J. Am. Med. Assoc.*, 273:535-541, 1995.
- Shi, Liu, Lippman, Dickson, *Human Reprod.*, Suppl. 1, 9:162-173, 1994.
- Simard *et al.*, "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families," *Nature Genet.*, 8:392-398, 1994.
- Smith and Johnson, "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase," *Gene*, 67:31-40, 1988.
- Soule *et al.*, "A human cell line from a pleural effusion derived from a breast carcinoma," *J. Nat. Cancer Inst.*, 51:1409-1413, 1973.
- Spillman and Bowcock, *Am. J. Hu. Gen.*, 57 Supplement, A5, 1995.
- Steck *et al.*, "Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers," *Nature Genet.* 15:356-362, 1997.
- Stratton, "Recent advances in understanding of genetic susceptibility to breast cancer." *Hum. Molec. Genet.*, 5:1515-1519, 1996.
- Takahashi *et al.*, "Mutations of the BRCA2 gene in ovarian carcinomas." *Cancer Res.*, 56:2738-2741, 1996.
- Tavtigian, Simard, Rommens, Couch, Shattuck-Eidens, Neuhausen, Merajver, Thorlacius, Offit, Stoppa-Lyonnet, Belanger, Bell, Berry, Bogden, Chen, Davis, Dumont, Frye, Hattier, Jammulapati, Janecki, Jiang, Kehrer, Leblanc, Mitchell, McArthur-Morrison, Nguyen, Peng, Samson, Schroeder, Snyder, Steele, Stringfellow, Stroup, Swedlund, Swensen, Teng, Thomas, Tran, Tran, Tranchant, Weaver-Feldhaus, Wong, Shizuya, Eyfjord, Cannon-Albright, Labrie, Skolnick, Weber, Kamb, Goldgar, *Nature Gen.*, 12:333-337, 1996.
- Telenius *et al.*, "Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer," *Genomics*, 13:718-725, 1992.
- Teng, "Low incidence of BRCA2 mutations in breast carcinoma and other cancers," *Nature Genet.*, 13:241-244, 1996.
- Thakur, Zhang, Peng, Le, Carroll, Ward, Yao, Farid, Couch, Wilson, Weber, *Mol. Cell. Biol.*, 17:444-452, 1997.

- Thomas, Smith, Rubinfeld, Gutowski, Beckmann, Polakis, *J. Biol. Chem.*, 271:28630-28635, 1996.
- Thompson *et al.*, "Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression," *Nature Genet.*, 9:444-450, 1995.
- 5 Trask, In: *Fluorescence in situ hybridization*, Birren *et al.*, eds., Cold Spring Harbor Laboratory Press, 1997.
- Vaughn, Davis, Jarboe, Huper, Evans, Wiseman, Berchuck, Inglehart, Futreal, Marks, *Cell Growth and Diff.*, 7:711-715, 1996a.
- Vaughn *et al.*, Cirisano, Huper, Berchuck, Futreal, Marks, Iglehart, *Cancer Res.*, 56:4590-4594, 10 1996b.
- Vogelstein and Kinzler, "Has the breast cancer susceptibility gene been found?" *Cell*, 79:1-3, 1994.
- Wakeling, Dukes, Bowler, *Cancer Res.*, 51:3867-3873, 1991.
- Wakeling, Newbould, Peters, *J. Molec. Endocrinol.*, 2:225-234, 1989.
- 15 Wang, Lin, Su, Hung, *Biochem. Biophys. Res. Comm.*, 234:247-251, 1997.
- Weber *et al.*, "A somatic truncating mutation in BRCA2 in a sporadic breast tumor," *Am. J. Hum. Genet.*, 59:962-964, 1996.
- Weinberger *et al.*, *Science*, 228:740-742, 1985.
- Wise *et al.*, "Identification and localization of the gene for EXTL, a third member of the 20 multiple exostoses gene family," *Genome Res.*, In Press, 1997.
- Wolf *et al.*, "An Integrated Family of Amino Acid Sequence Analysis Programs," *Comput. Appl. Biosci.*, 4(1):187-191, 1988.
- Wong *et al.*, "Appearance of  $\beta$ -lactamase activity in animal cells upon liposome mediated gene transfer," *Gene*, 10:87-94, 1980.
- 25 Wooster, Neuhausen, Mangion, Quirk, Ford, Collins, Nguyen, Seal, Tran, Averill, Fields, Marshall, Narod, Lenoir, Lynch, Feunteun, Devilee, Cornelisse, Menko, Daly, Ormiston, McManus, Pye, Lewis, Cannon-Albright, Peto, Ponder, Skolnick, Easton, Goldgar, Stratton, *Science*, 265:2088-2090, 1994.

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: Board of Regents, University of Texas System
- (B) STREET: 201 West 7th Street
- (C) CITY: Austin
- (D) STATE: TX
- (E) COUNTRY: US
- (F) POSTAL CODE (ZIP): 78701
- (G) TELEPHONE: (512) 418-3000
- (H) TELEFAX: (713) 789-2679

(ii) TITLE OF INVENTION: Compositions and Methods Comprising BARD1 and Other BRCA1 Binding Proteins

(iii) NUMBER OF SEQUENCES: 130

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/025,296
- (B) FILING DATE: 20-SEP-1996

## (vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/042,611
- (B) FILING DATE: 03-APR-1997

## (vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/042,985
- (B) FILING DATE: 04-APR-1997

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 531
- (D) OTHER INFORMATION: /note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 153
- (D) OTHER INFORMATION: /note= "Xaa = Glu or Lys for both  
SEQ ID NO:1 and SEQ ID NO:2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT	60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro	110
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro	158
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu	206
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro	254
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser	302
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile	350
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys	398
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys	446
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys	494
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr	542
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp	590
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro	638
GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys	782



CAA AAG CTG GTA TTC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406
GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550

CAA Gln	AAT Asn	GAC Asp 495	TCA Ser	CCA Pro	CTT Leu	CAC His	GAT Asp 500	GCA Ala	GCC Ala	AAG Lys	AAT Asn	GGG Gly 505	CAC His	GTG Val	GAT Asp	1598
ATA Ile	GTC Val 510	AAG Lys	CTG Leu	TTA Leu	CTT Leu	TCC Ser 515	TAT Tyr	GGA Gly	GCC Ala	TCC Ser	AGA Arg 520	AAT Asn	GCT Ala	GTT Val	AAT Asn	1646
ATA Ile 525	TTT Phe	GGT Gly	CTG Leu	CGG Arg	CCT Pro 530	GTC Val	GAT Asp	TAT Tyr	ACA Thr	GAT Asp 535	GAT Asp	GAA Glu	AGT Ser	ATG Met	AAA Lys 540	1694
TCG Ser	CTA Leu	TTG Leu	CTG Leu	CTA Leu 545	CCA Pro	GAG Glu	AAG Lys	AAT Asn	GAA Glu 550	TCA Ser	TCC Ser	TCA Ser	GCT Ala	AGC Ser 555	CAC His	1742
TGC Cys	TCA Ser	GTA Val 560	ATG Met	AAC Asn	ACT Thr	GGG Gly	CAG Gln	CGT Arg 565	AGG Arg	GAT Asp	GGA Gly	CCT Pro	CTT Leu 570	GTA Val	CTT Leu	1790
ATA Ile	GGC Gly 575	AGT Ser	GGG Gly	CTG Leu	TCT Ser	TCA Ser	GAA Glu 580	CAA Gln	CAG Gln	AAA Lys	ATG Met 585	CTC Leu	AGT Ser	GAG Glu	CTT Leu	1838
GCA Ala 590	GTA Val	ATT Ile	CTT Leu	AAG Lys	GCT Ala 595	AAA Lys	AAA Lys	TAT Tyr	ACT Thr	GAG Glu	TTT Phe 600	GAC Asp	AGT Ser	ACA Thr	GTA Val	1886
ACT Thr 605	CAT His	GTT Val	GTT Val	GTT Val	CCT Pro 610	GGT Gly	GAT Asp	GCA Ala	GTT Val	CAA Gln 615	AGT Ser	ACC Thr	TTG Leu	AAG Lys	TGT Cys 620	1934
ATG Met	CTT Leu	GGG Gly	ATT Ile	CTC Leu 625	AAT Asn	GGA Gly	TGC Cys	TGG Trp	ATT Ile 630	CTA Leu	AAA Lys	TTT Phe	GAA Glu	TGG Trp 635	GTA Val	1982
AAA Lys	GCA Ala	TGT Cys 640	CTA Leu	CGA Arg	AGA Arg	AAA Lys	GTA Val 645	TGT Cys 645	GAA Glu	CAG Gln	GAA Glu	GAA Glu	AAG Lys 650	TAT Tyr	GAA Glu	2030
ATT Ile	CCT Pro	GAA Glu 655	GGT Gly	CCA Pro	CGC Arg	AGA Arg	AGC Ser 660	AGG Arg	CTC Leu	AAC Asn	AGA Arg	GAA Glu 665	CAG Gln	CTG Leu	TTG Leu	2078
CCA Pro 670	AAG Lys	CTG Leu	TTT Phe	GAT Asp	GGA Gly 675	TGC Cys	TAC Tyr	TTC Phe	TAT Tyr	TTG Leu	TGG Trp 680	GGA Gly	ACC Thr	TTC Phe	AAA Lys	2126
CAC His 685	CAT His	CCA Pro	AAG Lys	GAC Asp	AAC Asn 690	CTT Leu	ATT Ile	AAG Lys	CTC Leu	GTC Val 695	ACT Thr	GCA Ala	GGT Gly	GGG Gly	GGC Gly 700	2174
CAG Gln	ATC Ile	CTC Leu	AGT Ser	AGA Arg 705	AAG Lys	CCC Pro	AAG Lys	CCA Pro	GAC Asp 710	AGT Ser	GAC Asp	GTG Val	ACT Thr	CAG Gln 715	ACC Thr	2222
ATC Ile	AAT Asn	ACA Thr 720	GTC Val	GCA Ala	TAC Tyr	CAT His	GCG Ala 725	AGA Arg	CCC Pro	GAT Asp	TCT Ser	GAT Asp	CAG Gln 730	CGC Arg	TTC Phe	2270
TGC Cys	ACA Thr 735	CAG Gln	TAT Tyr	ATC Ile	ATC Ile	TAT Tyr	GAA Glu 740	GAT Asp	TTG Leu	TGT Cys	AAT Asn	TAT Tyr 745	CAC His	CCA Pro	GAG Glu	2318

AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760

GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775

ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475

TTTTAATGTT CACATTTTAA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15

Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30

Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45

Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60

Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80

Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95

Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110

Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125

Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140

Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175

Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190

Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205

Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220

Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510  
 Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575

Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Ala Asp Tyr Lys Asp Asp Asp Asp Lys Ser  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTACCATGGA TTTATCTGCT CTTCGCGTT

## (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AAAAGTCGAC TAGAATTCAG CCTTTTCTAC ATTCATTC

38

## (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AACAGTACAA TGACTGGGCT C

21

## (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TCAGCGCTTC TGCACACAGT

20

## (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met	Ala	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Ser	Leu	Arg	Ser
1				5					10					15

## (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 993 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCTTCGTGGC CAGAAAGCAA AGTAACAGAA TTTCTCCATC AAAGTAAATT AAAATCTTTT

60

GAAAGTGAGC GTGTTCAACT TCTGCAAGAG GAAACAGCAA GAAATCTCAC ACAGTGTCAA	120
TTGGAATGTG AAAAATATCA GAAAAAATTG GAGGTTTTAA CCAAAGAATT TTATAGTCTC	180
CAAGCCTCTT CTGAAAAACG CATTACTGAA CTTCAAGCAC AGAACTCAGA GCATCAAGCA	240
AGGCTAGACA TTTATGAGAA ACTGGAAAAA GAGCTTGATG AAATAATAAT GCAAACTGCA	300
GAAATTGAAA ATGAAGATGA GGCTGAAAGG GTTCTTTTTT CCTACGGCTA TGGTGCTAAT	360
GTCCCCACAA CAGCCAAAAG ACGACTAAAG CAAAGTGTTT ACTTGGCAAG AAGAGTGCTT	420
CAATTAGAAA AACAAAATC GCTGATTTTA AAAGATCTGG AACATCGAAA GGACCAAGTA	480
ACACAGCTTT CACAAGAGCT TGACAGAGCC AATTCGCTAT TAAACCAGAC TCAACAGCCT	540
TACAGGTATC TCATTGAATC AGTGCGTCAG AGAGATTCTA AGATTGATTC ACTGACGGAA	600
TCTATTGCAC AACTTGAGAA AGATGTCAGC AACTTAAATA AAGAAAAGTC AGCTTTACTA	660
CAGACGAAGA ATCAAATGGC ATTAGATTTA GAACAACTTC TAAATCATCG TGAGGAATTG	720
GCAGCAATGA AACAGATTCT CGTTAAGATG CATAGTAAAC ATTCTGAGAA CAGCTTACTT	780
CTCACTAAAA CAGAACCAAA ACATGTGACA GAAAATCAGA AATCAAAGAC TTTGAATGTG	840
CCTAAAGAGC ATGAAGACAA TATATTTACA CCTAAACCAA CACTCTTTAC TAAAAAGAA	900
GCACCTGAGT GGTCTAAGAA ACAAAGATG AAGACCTAGT GTTTGGATG GGAAGCACCT	960
GTAGACCATT ATATACTCCT GAAGTTCTTT TTC	993

## (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1770 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CGTTCAAAGA GGGAGTTCAT TCAGGAACCT GCTAAGAATC GGCCCGGTCC CCAGACACGA	60
TCAGACCTAC TGCTGTCAGG AAGGGACTGG AATACGCTAA TTGTGGGAAA GCTTTCTCCA	120
TGGATTGCTC CAGACTCAA AGTGGAGAAG ATTGCAGGA ACTCCGAGGC GGCCATGTTA	180
CAGGAGCTGA ATTTTGGTGC ATATTTGGGT CTTCCAGCTT TCCTGCTGCC CCTTAATCAG	240
GAAGATAACA CCAACCTGGC CAGAGTTTTG ACCAACCACA TCCACACTGG CCATCACTCT	300
TCCATGTTCT GGATGCGGGT ACCCTTGGTG GCACCAGAGG ACCTGAGAGA TGATATAATT	360
GAGAATGCAC CAACTACACA CACAGAGGAG TACAGTGGGG AGGAGAAAAC GTGGATGTGG	420
TGGCACAACT TCCGGACTTT GTGTGACTAT AGTAAGAGGA TTGCAGTGGC TCTTGAAATT	480
GGGGCTGACC TCCCATCTAA TCATGTCATT GATCGCTGGC TTGGGGAGCC CATCAAAGGA	540
GGCATTCTCC CCACTAGCAT TTCCCTGACC AATAAGAAGG GATTCCTGT TCTTTCTAAG	600
ATGCACCAGA GGCTCATCTT CCGGCTCCTC AAGTTGGAGG TGCAGTTCAT CATCACAGGC	660
ACCAACCACC ACTCAGAGAA GGAGTTCTGC TCCTACCTCC AATACCTGGA ATACTTAAGC	720

CAGAACCGCC CTCCACCTAA TGCCTATGAA CTCTTTGACA AGGGCTATGA AGACTATCTG	780
CAGTCCCCGC TTCAGCCACT GATGGACAAT CTGGAATCTC AGACATATGA AGTGTTTGAA	840
AAGGACCCCA TCAAATACTC TCAGTACCAG CAGGCCATCT ATAAATGTCT GCTAGACCGA	900
GTACCAGAAG AGGAGAAGGA TACCAATGTC CAGGTAAGTA TGGTGCTGGG AGCAGGACGG	960
GGACCCCTGG TGAACGCTTC CCTGCGGGCA GCCAAGCAGG CCGACCGGCG GATAAAGCTG	1020
TATGCTGTGG AGAAAAACCC AAATGCCGTG GTGACGCTAG AGAACTGGCA GTTTGAAGAA	1080
TGGGGAAGCC AAGTGACCGT AGTCTCATCA GACATGAGGG AATGGGTGGC TCCAGAGAAA	1140
GCAGACATCA TTGTCAAGTA GCTTCTGGGC TCATTGTCTG ACAATGAATT GTCGCCTGAG	1200
TGCCTGGATG GAGCCCAGCA CTTCTTAAAA GATGATGGTG TGAGCATCCC CGGGGAGTAC	1260
ACTTCCTTTC TGGCTCCCAT CTCTTCCTCC AAGCTGTACA ATGAGGTCCG AGCCTGTAGG	1320
GAGAAGGACC GTGACCCTGA GGCCAGTTT GAGATGCCTT ATGTGGTACG GCTGCACAAC	1380
TTCCACCAGC TCTCTGCACC CCAGCCCTGT TTCACCTTCA GCCATCCCAA CAGAGATCCT	1440
ATGATTGACA ACAACCGCTA TTGCACCTTG GAATTTCTCTG TGGAGGTGAA CACAGTACTA	1500
CATGGCTTTG CCGGCTACTT TGAGACTGTG CTTTATCAGG ACATCACTCT GAGTATCCGT	1560
CCAGAGACTC ACTCTCTGG GATGTTCTCA TGGTTTCCCA TCCTCTTCCC TATTAAGCAG	1620
CCCATAACGG TACGTGAAGG CCAAACCATC TGTGTGCGTT TCTGGCGATG CAGCAATTCC	1680
AAGAAGGTGT GGTATGAGTG GGCTGTGACA GCACCAGTCT GTTCTGCTAT TCATAACCCC	1740
ACAGGCCGCT CATATACCAT TGGCCTCTAG	1770

## (2) INFORMATION FOR SEQ ID NO: 11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1345 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 328
- (D) OTHER INFORMATION: /note= "R = A or G"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GAGCCCGGCC GCGGCCTGCT GGTTTCAGTG ATGGCTCATG AAGCAATGGA ATATGATGTT	60
CAGGTGCAGT TAAATCATGC CGAACAACAG CCAGCTCCTG CTGGCATGGC CAGCAGCCAA	120
GGGGGACCAG CCCTCCTCCA GCCTGTTTCTT GCTGATGTGG TCAGCAGCCA GGGGGTACCA	180
TCCATCCTCC AGCCAGCTCC TGCTGAGGTG ATCAGCAGCC AAGCGACACC ACCCCTGCTC	240
CAGCCTGCTC CGCAACTGTC TGTGACCTG ACAGAAGTGG AGGTCTTGGG AGAAGACACT	300
GTGGAGAACA TCAATCCAAG AACTTCARAA CAACATAGGC AGGGATCTGA TGGTAATCAC	360
ACCATCCCAG CATCTTCGTT GCATTCAATG ACCAACTTCA TCAGCGGACT GCAGAGACTT	420



CATGGCATGC TGAATTCTT GAGACCTTCA TCTTCAAACC ACAGTGTAGG GCCAATGAGA	480
ACAAGAAGGA GGGTATCTGC TTCACGGAGG GCAAGAGCCG GAGGGTCTCA GAGGACAGAC	540
AGTGCCAGGT TGAGAGCACC ATTGGATGCT TACTTTCAGG TGAGCAGGAC CCAGCCTGAC	600
TTGCCAGCTA CCACTTATGA TTCAGAGACT AGGAATCCTG TATCTGAAGA GTTGCAGGTG	660
TCTAGTAGTT CTGATTCTGA CAGTGACAGC TCTGCAGAGT ATGGAGGGGT TGTGACCAC	720
GCAGAGGAAT CTGGAGCTGT CATTTTAGAA GAGCAACTAG CAGGTGTCTC AGCAGAGCAA	780
GAAGTTACAT GTATCGATGG AGGCAAGACC CTCCCCAAAC AGCCATCTCC CCAGAAGTCT	840
GAGCCTCTGC TACCTTCTGC TTCTATGGAT GAGGAAGAAG GGGACACTTG TACAATATGT	900
CTGGAACAGT GGACCAATGC TGGGGACCAC CGGCTCTCAG CATTACGCTG TGGGCATCTC	960
TTTGGGTATA GGTGCATTTT CACGTGGCTT AAAGGACAAG TACGAAAATG TCCCCAGTGC	1020
AACAAGAAAG CCAGGCACAG TGACATTGTC GTCCTTTATG CCCGAACCCT GAGAGCTTTG	1080
GACACTAGTG AACAGGAGCG CATGAAAAGG TAGGTGGTAA GAGTATGCCT GGCTGGAATG	1140
TTCCCTTTTG GTTCATTGTA GGCACATCTG AAAAAGAAGT TATGAGTCAC TCGTAGTGAG	1200
GTTTTACTTG ACCTGTGACT TGGGATCTCT GGGGATCATT GGCAGTCTGT CTTACACTGT	1260
TATTTATAAT TCATGTCTGA TCATCTCTT AAGGAAGTCT GCATCGTTTG CCTTATGTAG	1320
AGCATTAAAC ACAAGGATCT GGCAC	1345

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1248 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 786
- (D) OTHER INFORMATION: /note= "R = A or G"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GACTACCATC AGAACTGGGG CCGTGATGGG GGTCCCCGCA GCTCCGGTGG GGGCTATGGA	60
GGGGGGCCAG CAGGGGGTCA TGGAGGTAAC CGAGGCTCCG GAGGAGGCGG CGGCGGCGGA	120
GGGGGTGGTC GAGGCGGCAG GGGCCGGCAT CCCGGGCACC TGAAAGGCCG CGAAATCGGC	180
ATGTGGTACG CGAAAAACA GGGGCAGAAG AACAAGGAAG CGGAGAGGCA AGAGAGAGCT	240
GTA GTACACA TGGATGAACG ACGAGAAGAA CAAATTGTAC AGTTACTGAA TTCTGTTCAA	300
GCGAAGAATG ATAAAGAGTC AGAAGCACAG ATATCCTGGT TTGCTCCTGA GGATCATGGA	360
TACGGTACTG AAGTTTCTAC TAAGAACACA CCATGCTCAG AGAACAACT TGACATCCAG	420
GAAAGAAGT TGATAAATCA AGAAAAAAA ATGTTTAGAA TCAGGAACAG ATCATATATT	480
GACCGAGATT CTGAGTATCT CTTGCAAGAA AATGAACCAG ATGGAACTTT AGACCAAAAA	540

TTATTGGAAG	ATTACAAAA	GAAAAAAAT	GACCTTCGGT	ATATTGAAAT	GCAGCATTTTC	600
AGAGAAAAGC	TGCCTTCGTA	TGGAATGCAA	AAGGAATTGG	TAAATTTAAT	TGATAACCAT	660
CAGGTAACAG	TAATAAGTGG	TGAAACTGGT	TGTGGCAAAA	CCACTCAAGT	TACTCAGTTC	720
ATTTTGGATA	ACTACATTGA	AAGAGGAAAA	GGATCTGCTT	GCAGAATAGT	TTGTACTCAG	780
CCAAGRAGAA	TTAGTGCCAT	TTCAGTTGCG	GAAAGAGTAG	CTGCAGAAAAG	GGCAGAATCT	840
TGTGGCAGTG	GTAATAGTAC	TGGATATCAA	ATTTCGTCTCC	AGAGTCGGTT	GCCAAGGAAA	900
CAGGGTTCTA	TCTTATACTG	TACAACAGGA	ATCATCCTTC	AGTGGCTCCA	GTCAGACCCG	960
TATTTGTCCA	GTGTTAGTCA	TATCGTACTT	GATGAAATCC	ATGAAAGAAA	TCTGCAGTCA	1020
GATGTTTTAA	TGACTGTTGT	TAAAGACCTT	CTCAATTTTC	GATCTGACTT	GAAAGTAATA	1080
TTGATGAGTG	CAACATTGAA	TGCAGAAAAG	TTTTCAGAAT	ATTTTGGTAA	CTGTCCAATG	1140
ATACATATAC	CTGGTTTTAC	CTTTCCGGTT	GTGGAATATC	TTTTGGAAGA	TGTAATTGAA	1200
AAAATAAGGT	ATGTTCCAGA	ACAAAAAGAA	CACAGATCCC	AGTTTAAG		1248

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1803 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified base
- (B) LOCATION: 1362..1771
- (D) OTHER INFORMATION: /note= "N = A or C or G or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AATATATCCT	GGAAGAAGAC	AATAGTTACC	CGTTTCCTAA	AACTGGTTCC	AGACCTTTTG	60
GCCATTGTGC	AGCGTAAGAA	AAAGGAAGGG	GAAGAAGAAC	AAGCAATCAA	CAGACAGACA	120
GCGTTGTATA	CCTTAAAGCT	TTTATGCAAG	AATTTTGGTG	CAGAAAATCC	AGATCCTTTT	180
GTCCCAGTGC	TGAGCACTGC	TGTGAACTG	ATTGCTCCAG	AGAGAAAGGA	GGAGAAGAAT	240
GTCTTGGGAA	GCGCGCTGCT	GTGCATAGCA	GAGGTGACCT	CCACCCTGGA	GGCGCTGGCC	300
ATCCCCCAGC	TTCCCAGCCT	GATGCCATCG	TTGCTGACAA	CAATGAAGAA	CACCAGCGAG	360
CTGGTCTCCA	GCGAGGTCTA	CCTGCTCAGT	GCCTTGGCTG	CTCTGCAGAA	GGTTGTGGAG	420
ACTCTCCCGC	ACTTCATCAG	CCCCTATCTG	GAAGGCATTC	TCTCCAGGT	GATTCATCTG	480
GAGAAAATCA	CTAGTGAAAT	GGGTTCTGCG	TCACAGGCTA	ATATCCGCCT	CACATCTCTT	540
AAAAAGACAC	TGGCTACCAC	ACTTGACCCC	CGAGTCCTGT	TGCCCCGCAT	CAAAAAAACT	600
TACAAGCAGA	TTGAGAAGAA	CTGGAAGAAT	CACATGGGTC	CGTTTATGAG	CATCTTGCAA	660
GAGCATATTG	GGGCGATGAA	GAAGGAAGAG	CTCACCTCCC	ATCAGTCTCA	GCTAACCGCC	720
TTTTTCCTGG	AGGCCCTGGA	CTTCCGAGCC	CAGCACTCTG	AGAACGATCT	GGAGGAAGTT	780

GGAAAAACGG AAAATTGTAT CATTGACTGT CTAGTAGCCA TGGTTGTCAA ACTTTCCGAG	840
GTCACATTCA GGCCCCTGTT CTTCAAGCTG TTTGATTGGG CTAAACAGA AGATGCCCCA	900
AAGGACAGGT TGTTGACATT TTACAACCTG GCAGATTGCA TTGCTGAAAA GCTGAAAGGG	960
CTTTTACTC TGTTTGCCGG CCACTTAGTG AAGCCTTTTG CTGACACCTT GGACCAGGTG	1020
AACATCTCCA AAACAGATGA AGCATTTTTT GACTCTGAAA ATGACCCTGA AAAGTGCTGC	1080
TTGCTGTTGC AGTTTATTTT GAACTGTTTA TACAAAATCT TCCTTTTTGA TACCCAGCAT	1140
TTTATAAGTA AAGAGAGAGC AGGAGCCTTG ATGATGCCTC TGGTGGATCA GCTGGAAAAC	1200
AGGCTTGGGG GAGAAGAGAA ATTCCAGGAA CGGGTGACAA AGCACCTGAT ACCATGCATC	1260
GTACAGTTTT CCGTGCCAT GCGGATGAC TCTCTTTGGA AACCCTGAA CTACCAGATT	1320
CTGCTAAAGA CGAGAGACTC CTCGCTAAG GTTCGATTG NTGCTTTGAT TACTGTGTTA	1380
GCACTGGCTG AAAAATAAA GGAGAATTAT ATTGTCTTGC TACCAGAATC CATTCTTTTC	1440
TTAGCAGAGT TGATGGAAGA TGAATGTGAA GAAGTAGAAC ATCAGTGCCA AAAGACTATT	1500
CAGCAACTGG AAAGTGCCT GGGAGAGCCA CTCCAGAGCT ATTTCTAAGA CTTCTGTGGT	1560
GTTTCATACT CTACTCAGAG TTCACACTCA TATTTTCATAT TTTTATTTTC GGGTGTGGG	1620
TGCCATGTTA CTTTGGGTGT CTTAATACAC CTACTTGGAT TACTTACAAA TGTTTTATCA	1680
CTTCGNTACA AAATCCCCAC CTGGCTTGTG CTGNACATA AGCCTCTCCC GCCTATCGNA	1740
TAGAGCTTGT AGAGGCCTCG CGGCCTCGAN AGATCTATTG AATCGCTAGA TACTGAAAAA	1800
ACC	1803

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 817 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TGGGGGTCGT CCCTAACGGC CGCGACGCAG AGAGCGGTCA CTCCCTGGCC GAGGGGCAGG	60
CTCCTCACGG CCTCCCTGGG ACCCCAGGCG CGTCGGGAGG CGTCGTCCTC CAGCCCCGAG	120
GCCGGCGAAG GGCAGATCCG CCTCACAGAC AGTTGCGTCC AGAGGCTTTT GGAAATCACC	180
GAAGGTCAGA ATTCCTCAGG CTGCAAGTGG AGGGAGGTGG ATGCTCCGGA TTCCAATACA	240
AATTTTCACT GGATACAGTT ATCAACCCCG ACGACAGGGT ATTTGAACAG GGTGGGGCAA	300
GAGTGGTGGT TGA CTCTGAT AGCTTGGCCT TCGTGAAAGG GGCCAGGTG GACTTCAGCC	360
AAGAACTGAT CCGAAGCTCA TTTCAAGTGT TGAACAATCC TCAAGCACAG CAAGGCTGCT	420
CCTGTGGGTC ATCTTTCTCT ATCAAATTTT GATGTGATGA CTGGTGA CTG TGGGATTGTC	480
ACCAGTTGTA CCAATTTGAA GAACCTGGAA TTAGTAGAAT TCTAGAAGTT TACTTCTAAT	540

CATGTCCCTC	TCAATTTTAT	TTCCCGCAGT	CCAGGAGTGT	TATGTTTTGC	CACTATTATT	600
TTCAGAATGT	GAAGATTTTA	CTCTTGGCTT	AATTTTTCCT	TCCACTCAGT	GCTAAGGCTG	660
AGCCTCCAGA	TGCTGTTACC	TCAGATTTAA	TCAGTGGTGT	AAACTCCGTA	TAATCTGTAG	720
AGCCTCCATG	GCTCTAAAAT	TTGGAATTAA	CTTCTCTTGC	CTTAAGAGCT	GCTTGTACAT	780
ATGTGGATAG	CTATGTATAA	AAGCTTCATT	TTAAAAA			817

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2138 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGCCCGGCTG	GAGGAGCCCC	GACCCAGCT	CTGGTGGCGG	GCAGCAGCGC	CGCGGCCCCC	60
TTCCCTCACG	GGGACTCGGC	CCTGAACGAG	CAGGAGAAGG	AGTTGCAGCG	GCGGCTGAAG	120
CGCCTCTACC	CGGCCGTGGA	CGAACAAGAG	ACGCCGCTGC	CTCGGTCCTG	GAGCCCGAAG	180
GACAAGTTCA	GCTACATCGG	CCTCTCTCAG	AACAACCTGC	GGGTGCACTA	CAAAGGTCAT	240
GGCAAAACCC	CAAAGATGC	CGCGTCAGTT	CGAGCCACGC	ATCCAATACC	AGCAGCCTGT	300
GGGATTTATT	ATTTTGAAGT	AAAAATTGTC	AGTAAGGGAA	GAGATGGTTA	CATGGGAATT	360
GGTCTTTCTG	CTCAAGGTGT	GAACATGAAT	AGACTACCAG	GTTGGGATAA	GCATTCATAT	420
GGTTACCATG	GGGATGATGG	ACATTCGTTT	TGTTCTTCTG	GAAGTGGACA	ACCTTATGGA	480
CCAACTTTCA	CTACTGGTGA	TGTCATTGGC	TGTTGTGTTA	ATCTTATCAA	CAATACCTGC	540
TTTTACACCA	AGAATGGACA	TAGTTTAGGT	ATTGCTTTCA	CTGACCTACC	GCCAAATTTG	600
TATCCTACTG	TGGGGCTTCA	AACACCAGGA	GAAGTGGTCG	ATGCCAATTT	TGGGCAACAT	660
CCTTTCGTGT	TTGATATAGA	AGACTATATG	CGGGAGTGGA	GAACCAAAAT	CCAGGCACAG	720
ATAGATCGAT	TTCCTATCGG	AGATCGAGAA	GGAGAATGGC	AGACCATGAT	ACAAAAAATG	780
GTTTCATCTT	ATTTAGTCCA	CCATGGGTAC	TGTGCCACAG	CAGAGGCCTT	TGCCAGATCT	840
ACAGACCAGA	CCGTTCTAGA	AGAATTAGCT	TCCATTAAGA	ATAGACAAAG	AATTCAGAAA	900
TTGGTATTAG	CAGGAAGAAT	GGGAGAAGCC	ATTGAAACAA	CACAACAGTT	ATACCCAAGT	960
TTACTTGAAA	GAAATCCTAA	TCTCCTTTTC	ACATTAAAAG	TGCGTCAGTT	TATAGAAATG	1020
GTGAATGGTA	CAGATAGTGA	AGTACGATGT	TTGGGAGGCC	GAAGTCCAAA	GTCTCAAGAC	1080
AGTTATCCTG	TTAGTCCTCG	ACCTTTTAGT	AGTCCAAGTA	TGAGCCCCAG	CCATGGAATG	1140
AATATCCACA	ATTTAGCATC	AGGCAAAGGA	AGCACCGCAC	ATTTTTCAGG	TTTTGAAAGT	1200
TGTAGTAATG	GTGTAATATC	AAATAAGCA	CATCAATCAT	ATTGCCATAG	TAATAAACAC	1260
CAGTCATCCA	ACTTGAATGT	ACCAGAACTA	AACAGTATAA	ATATGTCAAG	ATCAGACAA	1320
GTTAATAACT	TCACCAGTAA	TGATGTAGAC	ATGGAAACAG	ATCACTACTC	CAATGGAGTT	1380

GGAGAACTT CATCCAATGG TTTCTTAAAT GGTAGCTCTA AACATGACCA CGAAATGGAA	1440
GATTGTGACA CCGAAATGGA AGTTGATTCA AGTCAGTTGA GACGCCAGTT GTGTGGAGGA	1500
AGTCAGGCCG CCATAGAAAG AATGATCCAC TTTGGACGAG AGCTGCAAGC AATGAGTGAA	1560
CAGCTAAGGA GAGACTGTGG CAAGAACACT GCAAACAAAA AATGTTGAAG GATGCATTCA	1620
GTCTACTAGC ATATTCAGAT CCTTGAACA GCCCAGTTGG AAATCAGCTT GACCCGATTC	1680
AGAGAGAACC TGTGTGCTCA GCTCTTAACA GTGCAATATT AGAAACCCAC AATCTGCCAA	1740
AGCAACCTCC ACTTGCCCTA GCAATGGGAC AGGCCACACA ATGTCTAGGA CTGATGGCTC	1800
GATCAGGAAT TGGATCCTGC GCATTTGCCA CAGTGAAGA CTACCTACAT TAGCTATGCA	1860
TTTCAAGAGC TCACACTTAT ATTGTGGCAT ATAGTCAACA TGAAGTAGA CCAGCTCTGC	1920
TGATTTGAAA TTTAGATTTT TTAAATTATG TACTGGGGAC AGGTTTTTGT CGCTTTACAT	1980
TGCTTCCTAG TTTACAGCAT GATGCAATG ATTTTCTAAC TTAGTGTTAG GAGAAATTAT	2040
TTTCCATCTT TAACCTCTTA GTTGTCTAAG AGTTAAATAT TACTGAATTT CAGACGTTCA	2100
AATTGATCAT CACAAATCCT TTAAACAAT TACCTAAA	2138

## (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3428 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION:178
  - (D) OTHER INFORMATION:/note= "W = A or T"
- (ix) FEATURE:
  - (A) NAME/KEY: modified base
  - (B) LOCATION:1331..3246
  - (D) OTHER INFORMATION:/note= "Y = C or T"
- (ix) FEATURE:
  - (A) NAME/KEY: modified base
  - (B) LOCATION:2886..3212
  - (D) OTHER INFORMATION:/note= "H = A or C or T"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGGCCGCCCC GCGGGAAGAT GAATAAGGGC TGGCTGGAGC TGGAGAGCGA CCCAGGCCTC	60
TTCACCCTGC TCGTGGAAGA TTTCGGTGTC AAGGGGGTGC AAGTGGAGGA GATCTACGAC	120
CTTCAGAGCA AATGTCAGGG CCCTGTATAT GGATTTATCT TCCTGTTCAA ATGGATCWAA	180
GAGCGCCGGT CCCGGCGAAA GGTCTCTACC TTGGTGGATG ATACGTCCGT GATTGATGAT	240
GATATTGTGA ATAACATGTT CTTTGCCAC CAGCTGATAC CCAACTCTTG TGCAACTCAT	300
GCCTTGCTGA GCGTGCTCCT GAACTGCAGC AGCGTGGACC TGGGACCCAC CCTGAGTCGC	360
ATGAAGGACT TCACCAAGGG TTTCAGCCCT GAGGCCCGAG CCACGCCACC TCCCTGAGAA	420

GCAGAATGGC CTTAGTGCAG TGCGGACCAT GGAGGCGTTC CACTTTGTCA GCTATGTGCC	480
TATCACAGGC CGGCTCTTTG AGCTGGATGG GCTGAAGGTC TACCCCATTG ACCATGGGCC	540
CTGGGGGGAG GACGAGGAGT GGACAGACAA GGCCCGGCGG GTCATCATGG AGCGTATCGG	600
CCTCGCCACT GCAGGGGAGC CCTACCACGA CATCCGCTTC AACCTGATGG CAGTGGTGCC	660
CGACCGCAGG ATCAAGTATG AGGCCAGGCT GCATGTGCTG AAGGTGAACC GTCAGACAGT	720
ACTAGAGGCT CTGCAGCAGC TGATAAGAGT AACAGAGCCA GAGCTGATTC AGACCCACAA	780
GTCTCAAGAG TCACAGCTGC CTGAGGAGTC CAAGTCAGCC AGCAACAAGT CCCCCTGGT	840
GCTGGAAGCA AACAGGGCCC CTGCAGCCTC TGAGGGCAAC CACACAGATG GTGCAGAGGA	900
GGCGGCTGGT TCATGCGCAC AAGCCCCATC CCACAGCCCT CCCAACAAAC CCAAGCTAGT	960
GGTGAAGCCT CCAGGCAGCA GCCTCAATGG GGTTCACCCC AACCCCACTC CCATTGTCCA	1020
GCGGCTGCCG GCCTTTCTAG ACAATCACAA TTATGCCAAG TCCCCATGC AGGAGGAAGA	1080
AGACCTGGCG GCAGGTGTGG GCCGCAGCCG AGTTCCAGTC CGCCACCCC AGCAGTACTC	1140
AGATGATGAG GATGACTATG AGGATGACGA GGAGGATGAC GTGCAGAACA CCAACTCTGC	1200
CCTTAGGTAT AAGGGGAAGG GAACAGGGAA GCCAGGGGCA TTGAGCGGTT CTGCTGATGG	1260
GCAACTGTCA GTGCTGCAGC CCAACACCAT CAACGTCTTG GCTGAGAAGC TCAAAGAGTC	1320
CCAGAAGGAC YTCTCAATTC CTCTGTCCAT CAAGACTAGC AGCGGGGCTG GGAGTCCGGC	1380
TGTGGCAGTG CCCACACACT CGCAGCCCTC ACCCACCCTC AGCAATGAGA GTACAGACAC	1440
GGCCTCTGAG ATCGGCAGTG CTTTCAACTC GCCACTGCGC TCGCCTATCC GCTCAGCCAA	1500
CCCGACGCGG CCCTCCAGCC CTGTACCTC CCACATCTCC AAGGTGCTTT TTGGAGAGGA	1560
TGACAGCCTG CTGCGTGTG ACTGCATACG CTACAACCGT GCTGTCCGTG ATCTGGGTCC	1620
TGTCATCAGC ACAGGCCTGC TGCACCTGGC TGAGGATGGG GTGCTGAGTC CCCTGGCGCT	1680
GACAGAGGGT GGGAAGGGTT CCTCGCCCTC CATCAGACCA ATCCAAGGCA GCCAGGGGTC	1740
CAGCAGCCCA GTGGAGAAGG AGGTCGTGGA AGCCACGGAC AGCAGAGAGA AGACGGGGAT	1800
GGTGAGGCCT GCGGAGCCCT TGAGTGGGGA GAAATACTCA CCCAAGGAGC TGCTGGCACT	1860
GCTGAAGTGT GTGGAGGCTG AGATTGCAAA CTATGAGGCG TGCCTCAAGG AGGAGGTAGA	1920
GAAGAGGAAG AAGTTCAAGA TTGATGACCA GAGAAGGACC CACAACCTACG ATGAGTTCAT	1980
CTGCACCTTT ATCTCCATGC TGGCTCAGGA AGGCATGCTG GCCAACCTAG TGGAGCAGAA	2040
CATCTCCGTG CGGCGGCGCC AAGGGGTCAG CATCGGCCGG CTCCACAAGC AGCGGAAGCC	2100
TGACCGGCGG AAACGCTCTC GCCCCTACAA GGCCAAGCGC CAGTGAGGAC TGCTGGCCCT	2160
GACTCTGCAG CCCACTCTTG CCGTGTGGCC CTCACCAGGG TCCTTCCCTG CCCCATTCC	2220
CCTTTTCCCA GTATTACTGA ATAGTCCCAG CTGGAGAGTC CAGGCCCTGG GAATGGGAGG	2280
AACCAGGCCA CATTCTTCC ATCGTGCCCT GAGGCCTGAC ACGGCAGATC AGCCCCATAG	2340
TGCTCAGGAG GCAGCATCTG GAGTTGGGGC ACAGCGAGGT ACTGCAGCTT CCTCCACAGC	2400

CGGCTGTGGA GCAGCAGGAC CTGGCCCTTC TGCCTGGGCA GCAGAATATA TATTTTACCT	2460
ATCAGAGACA TCTATTTTTC TGGGCTCCAA CCCAACATGC CACCATGTTG ACATAAGTTC	2520
CTACCTGACT ATGCTTTTCTC TCCTAAGGAG CTGTCCTGGT GGGCCCAGGT CCTTGTATCA	2580
TGCCACGGTC CCAACTACAG GGTCTAGCT GGGGGCCTGG GTGGGCCCTG GSCTCTGGGC	2640
CCTGCTGCTC TAGCCCCAGC CACCAGCCTG TCCCTGTTGT AAGGAAGCCA GGTCTTCTCT	2700
CTTCATTCTT CTTAGGAGAG TGCCAACTC AGGGACCCAG CACTGGGCTG GGTGGGAGT	2760
AGGGTGTCCC AGTGGGGTTG GGGTGAGCAG GCTGCTGGGA TCCCATGGCC TGAGCAGAGC	2820
ATGTGGGAAC TGTTCACTGG CCTGTGAAC GTCTTCCTTG TTCTAGCCAG GCTGTTCAAG	2880
ACTGCTHTCC ATAGCAAGGT TCTAGGGCTC TCGCCTTCA GTGTTGTGGC CCTAGCTATG	2940
GGCCTAAATT GGGCTCTAGG TCTCTGTCCC TGGCGCTTGA GGCTCAGAAG AGCCTCTGTC	3000
CAGCCCCCTCA GTATTACCAT GTCTCCCTCT CAGGGGTAGC AGAGACAGGG TTGCTTATAG	3060
GAAGCTGGCA CCACTCAGCT HTTCCTGCTA CTCCAGTTTC CTCAGCCTYT GCAAGGCACT	3120
CAGGGTGGGG GACAGCAGGA TCAAGACAAC CCGTTGGAGC CCCTGTGTTC CAGAGGACCT	3180
GATGCCAAGG GGTAATGGGC CCAGCAGTGC CHTTGAGCC CAGGCCCCAA CACAGCCCCA	3240
TGGCCTYTGC CAGATGGCTT TGAAAAGGT GATCCAAGCA GGCCCCCTTA TCTGTACATA	3300
GTGACTGAGT GGGGGGTGCT GGCAAGTGTG GCAGCTGCCT CTGGGCTGAG CACAGCTTGA	3360
CCCCTCTAGC CCCTGTAAAT ACTGGATCAA TGAATGAATA AAACCTCTCT AAGAATCTCC	3420
TGAAAAAA	3428

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 938 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAGCGGGAAG CAAGGGACGA GCTACCTGGA GCGCCTCCTG TTCTTTGCAG TTCCTCCTCA	60
GATCTTAGCC TCCTGTTGGG CCCCTCTTTT CAGAGCCAGC ATTCTTTCCA GCCCCTGGAG	120
CCCAAACCAG ACCTCACTTC ATCCACAGCT GGGGCCTTCT CTGCACTTGG GGCCTTCCAT	180
CCCGATCATA GGGCAGAAAG GCCATTCCCT GAGGAAGATC CTGGACCTGA CGGGGAGGGC	240
CTCCTAAAGC AAGGGCTGCC GCCTGCTCAG CTGGAGGGCC TCAAGAATTT TTTGCACCAG	300
TTGCTGGAGA CAGTGCCCCA GAACAATGAG AACCCTTCTG TCGACCTGTT GCCCCCTAAG	360
TCTGGTCCTC TGA CTGTCCC ATCTTGGGAG GAAGCCCCTC AAGTGCCACG TATTCCACCG	420
CCTGTCCACA AAACCAAAGT TCCCTTAGCC ATGGCATCCA GTCTTTTCCG GGTCCCTGAG	480
CCTCCCTCCT CCCATTCA CA AGGCAGTGGT CCCAGCAGTG GTTCCCCAGA GAGAGGTGGA	540

GATGGGCTTA CATTCCCAAG GCAGCTGATG GAGGTGTCTC AACTGTTGCG ACTCTACCAG 600  
GCTCGGGGCT GGGGGGCTCT GCCTGCTGAG GATCTCCTGC TCTACCTGAA GAGGCTGGAA 660  
CACAGCGGGA CTGATGGCCG AGGGGATAAT GTCCCCAGAA GGAACACAGA CTCCCGCTTG 720  
GGTGAGATCC CCCGGAAGA GATTCCCTCC CAGGCTGTCC CTCGCCGCCT TGCTACAGCC 780  
CCCAAGACTG AAAAACCTCC CGCACGGAAG AAAAGTGGGC ACCCTGCCCC GAGTAGCATG 840  
AGGAGCCGGG GGGGAGTCTG GAGATGAGCC CCCCTACCCT CTCTCCTCTT TGTTCTCTCA 900  
TTGTTGTTAT TTTAATAAAT GCTCAGTAGT CTGTAAAA 938

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1379 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ATGGTGAAGG TGAAGGGGCA GGTCAGCGAG ATGGCGGTGC TGCTCATCGA CCCCAGCCCT 60  
CAGATTGCTG CCCTGGCCAA GAACTTCTTC AATGAGCTCT CCCACAAGGG CAACGCAATC 120  
TATAATCTCC TTCCAGATAT CATCAGCCGC CTGTCAGACC CCGAGCTGGG GGTGGAGGAA 180  
GAGCCTTTCC ACACCATCAT GAAACAGCTC CTCTCCTACA TCACCAAGGA CAAGCAGACA 240  
GAGAGCCTGG TGGAAAAGCT GTGTCAGCGG TTCCGCACAT CCCTAACTGA GCGGCAGCAG 300  
CGAGACCTGG CCTACTGTGT GTCACAGCTG CCCCTCACAG AGCGAGGCCT CCGTAAGATG 360  
CTTGACAATT TTGACTGTTT TGGAGACAAA CTGTCAGATG AGTCCATCTT CAGTGCTTTT 420  
TTGTCAGTTG TAGGCAAGCT GCGACGTGGG GCCAAGCCTG AGGGCAAGGC TATAATAGAT 480  
GAATTTGAGC AGAAGCTTCG GGCCTGTCAT ACCAGAGGTT TGGATGGAAT CAAGGAGCTT 540  
GAGATTGGCC AAGCAGGTAG CCAGAGAGCG CCATCAGCCA AGAAACCATC CACTGGTTCT 600  
AGGTACCAGC CTCTGGCTTC TACAGCCTCA GACAATGACT TTGTCACACC AGAGCCCCGC 660  
CGTACTACCC GTCGGCATCC AAACACCCAG CAGCGAGCTT CCAAAAAGAA ACCCAAAGTT 720  
GTCTTCTCAA GTGATGAGTC CAGTGAGGAA GATCTTTCAG CAGAGATGAC AGAAGACGAG 780  
ACACCCAAGA AAACAACTCC CATTCTCAGA GCATCGGCTC GCAGGCACAG ATCCTAGGAA 840  
GTCTGTTCTT GTCCTCCCTG TGCAGGGTAT CCTGTAGGGT GACCTGGAAT TCGAATTCTG 900  
TTTCCCTTGT AAAATATTTG TCTGTCTCTT TTTTTTAAAA AAAAAAAGG CCGGGCACTG 960  
TGGCTCACGC CTGTAATCCC AGCACTTTGC GATACCAAGG CGGGTGGATA ACCTGAGGTA 1020  
GGGAGTTCGA GACCAGCCTG ACCAACATGG AGAAACCCCA TCTCTACTAA AAATAAAAAA 1080  
TTAGCCGGGC GTATTGGCGT GCGCCTGTAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG 1140  
AATCGCCTGA ACCCAGAGGC GGAGGTTGTA GTGAGCCGAA ATCACACCAT TGCACTCCAG 1200  
CTTGGGCAAC AATAGCGAAC CTCCATCTCA AATTAAAAA AAAATGCCTA CACGCTCTTT 1260



AAAATGCAAG GCTTTCTCTT AAATTAGCCT AACTGAACTG CGTTGAGCTG CTTCAACTTT 1320  
 GGAATATATG TTTGCCAATC TCCTTGTTTT CTAATGAATA AATGTTTTTA TATACTTTT 1379

## (2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 273 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Cys Ser Ser Ser Ser Asp Leu Ser Leu Leu Leu Gly Pro Ser Phe Gln  
 1 5 10 15  
 Ser Gln His Ser Phe Gln Pro Leu Glu Pro Lys Pro Asp Leu Thr Ser  
 20 25 30  
 Ser Thr Ala Gly Ala Phe Ser Ala Leu Gly Ala Phe His Pro Asp His  
 35 40 45  
 Arg Ala Glu Arg Pro Phe Pro Glu Glu Asp Pro Gly Pro Asp Gly Glu  
 50 55 60  
 Gly Leu Leu Lys Gln Gly Leu Pro Pro Ala Gln Leu Glu Gly Leu Lys  
 65 70 75 80  
 Asn Phe Leu His Gln Leu Leu Glu Thr Val Pro Gln Asn Asn Glu Asn  
 85 90 95  
 Pro Ser Val Asp Leu Leu Pro Pro Lys Ser Gly Pro Leu Thr Val Pro  
 100 105 110  
 Ser Trp Glu Glu Ala Pro Gln Val Pro Arg Ile Pro Pro Pro Val His  
 115 120 125  
 Lys Thr Lys Val Pro Leu Ala Met Ala Ser Ser Leu Phe Arg Val Pro  
 130 135 140  
 Glu Pro Pro Ser Ser His Ser Gln Gly Ser Gly Pro Ser Ser Gly Ser  
 145 150 155 160  
 Pro Glu Arg Gly Gly Asp Gly Leu Thr Phe Pro Arg Gln Leu Met Glu  
 165 170 175  
 Val Ser Gln Leu Leu Arg Leu Tyr Gln Ala Arg Gly Trp Gly Ala Leu  
 180 185 190  
 Pro Ala Glu Asp Leu Leu Leu Tyr Leu Lys Arg Leu Glu His Ser Gly  
 195 200 205  
 Thr Asp Gly Arg Gly Asp Asn Val Pro Arg Arg Asn Thr Asp Ser Arg  
 210 215 220  
 Leu Gly Glu Ile Pro Arg Lys Glu Ile Pro Ser Gln Ala Val Pro Arg  
 225 230 235 240  
 Arg Leu Ala Thr Ala Pro Lys Thr Glu Lys Pro Pro Ala Arg Lys Lys  
 245 250 255

Ser Gly His Pro Ala Pro Ser Ser Met Arg Ser Arg Gly Gly Val Trp  
 260 265 270

Arg

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 531
- (D) OTHER INFORMATION: /note= "R = A or G"

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 153
- (D) OTHER INFORMATION: /note= "Xaa = Glu or Lys for both  
 SEQ ID NO: 20 and SEQ ID NO: 21"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG  ATG  CCG  GAT  AAT  CGG  CAG  CCG  AGG  AAC  CGG  CAG  CCG      110
                Met  Pro  Asp  Asn  Arg  Gln  Pro  Arg  Asn  Arg  Gln  Pro
                1          5          10
AGG  ATC  CGC  TCC  GGG  AAC  GAG  CCT  CGT  TCC  GCG  TCC  GCC  ATG  GAA  CCG      158
Arg  Ile  Arg  Ser  Gly  Asn  Glu  Pro  Arg  Ser  Ala  Ser  Ala  Met  Glu  Pro
                15          20          25
GAT  GGT  CGC  GGT  GCC  TGG  GCC  CAC  AGT  CGC  GCC  GCG  CTC  GAC  CGC  CTG      206
Asp  Gly  Arg  Gly  Ala  Trp  Ala  His  Ser  Arg  Ala  Ala  Leu  Asp  Arg  Leu
                30          35          40
GAG  AAG  CTG  CTG  CGC  TGC  TCG  CGT  TGT  ACT  AAC  ATT  CTG  AGA  GAG  CCT      254
Glu  Lys  Leu  Leu  Arg  Cys  Ser  Arg  Cys  Thr  Asn  Ile  Leu  Arg  Glu  Pro
                45          50          55          60
GTG  TGT  TTA  GGA  GGA  TGT  GAG  CAC  ATC  TTC  TGT  AGT  AAT  TGT  GTA  AGT      302
Val  Cys  Leu  Gly  Gly  Cys  Glu  His  Ile  Phe  Cys  Ser  Asn  Cys  Val  Ser
                65          70          75
GAC  TGC  ATT  GGA  ACT  GGA  TGT  CCA  GTG  TGT  TAC  ACC  CCG  GCC  TGG  ATA      350
Asp  Cys  Ile  Gly  Thr  Gly  Cys  Pro  Val  Cys  Tyr  Thr  Pro  Ala  Trp  Ile
                80          85          90
CAA  GAC  TTG  AAG  ATA  AAT  AGA  CAA  CTG  GAC  AGC  ATG  ATT  CAA  CTT  TGT      398
Gln  Asp  Leu  Lys  Ile  Asn  Arg  Gln  Leu  Asp  Ser  Met  Ile  Gln  Leu  Cys
                95          100          105
AGT  AAG  CTT  CGA  AAT  TTG  CTA  CAT  GAC  AAT  GAG  CTG  TCA  GAT  TTG  AAA      446
Ser  Lys  Leu  Arg  Asn  Leu  Leu  His  Asp  Asn  Glu  Leu  Ser  Asp  Leu  Lys
                110          115          120

```

GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys 125 130 135 140	494
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr 145 150 155	542
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp 160 165 170	590
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro 175 180 185	638
GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214

AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT	1262
Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro	
385 390 395	
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA	1310
Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser	
400 405 410	
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA	1358
Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly	
415 420 425	
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT	1406
Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val	
430 435 440	
GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT	1454
Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala	
445 450 455 460	
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA	1502
Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val	
465 470 475	
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT	1550
Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr	
480 485 490	
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT	1598
Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp	
495 500 505	
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT	1646
Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn	
510 515 520	
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA	1694
Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys	
525 530 535 540	
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC	1742
Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His	
545 550 555	
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT	1790
Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu	
560 565 570	
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT	1838
Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu	
575 580 585	
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA	1886
Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val	
590 595 600	
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT	1934
Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys	
605 610 615 620	
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA	1982
Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val	
625 630 635	

AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174
CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr 705 710 715	2222
ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe 720 725 730	2270
TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu 735 740 745	2318
AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile 750 755 760	2366
GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser 765 770 775	2415
ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT	2475
TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC	2531

## (2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
1 5 10 15

Gly Asn Glu Pro Arg Ser Ala Ser Ala Met Glu Pro Asp Gly Arg Gly  
20 25 30

Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
35 40 45

Arg Cys Ser Arg Cys Thr Asn Ile L u Arg Glu Pro Val Cys Leu Gly  
50 55 60

Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
65 70 75 80

Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160  
 Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430

His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510  
 Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG      110
      Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro
              1              5              10
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG      158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro
      15              20              25
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG      206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu
      30              35              40
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT      254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro
      45              50              55              60
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT      302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser
              65              70              75
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA      350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile
              80              85              90
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT      398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys
              95              100              105
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA      446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys
      110              115              120
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG      494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys
      125              130              135              140
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG GAA GTC AGA TAT      542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Glu Val Arg Tyr
              145              150              155
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT      590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp
              160              165              170
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT      638
Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro
      175              180              185

```



GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTAA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Glu Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:24 and SEQ ID NO:25"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG  ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG      110
           Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro
           1             5             10
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG      158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro
           15             20             25
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG      206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu
           30             35             40
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT      254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro
           45             50             55             60
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT      302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser
           65             70             75
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA      350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile
           80             85             90
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT      398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys
           95             100             105
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA      446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys
           110             115             120
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG      494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys
           125             130             135             140
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT      542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr
           145             150             155
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT      590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp
           160             165             170
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT      638
Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro
           175             180             185

```

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACG GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174



CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2510 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ix) FEATURE:
- (A) NAME/KEY: CDS
  - (B) LOCATION: 75..2384

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:26 and SEQ ID NO:27"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG      110
      Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro
      1          5          10
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG      158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro
      15          20          25
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG      206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu
      30          35          40
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT      254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro
      45          50          55          60
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT      302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser
      65          70          75
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA      350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile
      80          85          90
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT      398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys
      95          100          105
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA      446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys
      110          115          120
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG      494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys
      125          130          135          140
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT      542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr
      145          150          155
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT      590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp
      160          165          170
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT      638
Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro
      175          180          185

```

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA CCT TCA TGC AAA CGT AAA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Pro Ser Cys Lys Arg Lys 350 355 360	1166
GTT GGT GGT ACA TCA GGG AGG AAA AAC AGT AAC ATG TCC GAT GAA TTC Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser Asp Glu Phe 365 370 375 380	1214
ATT AGT CTT TCA CCA GGT ACA CCA CCT TCT ACA TTA AGT AGT TCA AGT Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser Ser Ser Ser 385 390 395	1262
TAC AGG CAA GTG ATG TCT AGT CCC TCA GCA ATG AAG CTG TTG CCC AAT Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu Leu Pro Asn 400 405 410	1310
ATG GCT GTG AAA AGA AAT CAT AGA GGA GAG ACT TTG CTC CAT ATT GCT Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu His Ile Ala 415 420 425	1358
TCT ATT AAG GGC GAC ATA CCT TCT GTT GAA TAC CTT TTA CAA AAT GGA Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu Gln Asn Gly 430 435 440	1406

AGT Ser 445	GAT Asp 445	CCA Pro	AAT Asn	GTT Val	AAA Lys 450	GAC Asp	CAT His	GCT Ala	GGA Gly 455	TGG Trp 455	ACA Thr	CCA Pro	TTG Leu	CAT His	GAA Glu 460	1454
GCT Ala	TGC Cys	AAT Asn	CAT His	GGG Gly 465	CAC His	CTG Leu	AAG Lys	GTA Val	GTG Val 470	GAA Glu	TTA Leu	TTG Leu	CTC Leu	CAG Gln 475	CAT His	1502
AAG Lys	GCA Ala	TTG Leu	GTG Val 480	AAC Asn	ACC Thr	ACC Thr	GGG Gly	TAT Tyr 485	CAA Gln	AAT Asn	GAC Asp	TCA Ser	CCA Pro	CTT Leu	CAC His	1550
GAT Asp	GCA Ala	GCC Ala	AAG Lys 495	AAT Asn	GGG Gly	CAC His	GTG Val 500	GAT Asp	ATA Ile	GTC Val	AAG Lys	CTG Leu 505	TTA Leu	CTT Leu	TCC Ser	1598
TAT Tyr 510	GGA Gly 510	GCC Ala	TCC Ser	AGA Arg	AAT Asn	GCT Ala 515	GTT Val	AAT Asn	ATA Ile	TTT Phe	GGT Gly 520	CTG Leu	CGG Arg	CCT Pro	GTC Val	1646
GAT Asp 525	TAT Tyr	ACA Thr	GAT Asp	GAT Asp	GAA Glu 530	AGT Ser	ATG Met	AAA Lys	TCG Ser	CTA Leu 535	TTG Leu	CTG Leu	CTA Leu	CCA Pro	GAG Glu 540	1694
AAG Lys	AAT Asn	GAA Glu	TCA Ser	TCC Ser 545	TCA Ser	GCT Ala	AGC Ser	CAC His	TGC Cys 550	TCA Ser	GTA Val	ATG Met	AAC Asn	ACT Thr 555	GGG Gly	1742
CAG Gln	CGT Arg	AGG Arg	GAT Asp 560	GGA Gly	CCT Pro	CTT Leu	GTA Val	CTT Leu	ATA Ile	GGC Gly	AGT Ser	GGG Gly	CTG Leu	TCT Ser	TCA Ser	1790
GAA Glu	CAA Gln	CAG Gln 575	AAA Lys	ATG Met	CTC Leu	AGT Ser	GAG Glu	CTT Leu	GCA Ala	GTA Val	ATT Ile	CTT Leu	AAG Lys	GCT Ala	AAA Lys	1838
AAA Lys 590	TAT Tyr	ACT Thr	GAG Glu	TTT Phe	GAC Asp	AGT Ser 595	ACA Thr	GTA Val	ACT Thr	CAT His	GTT Val 600	GTT Val	GTT Val	CCT Pro	GGT Gly	1886
GAT Asp 605	GCA Ala	GTT Val	CAA Gln	AGT Ser	ACC Thr 610	TTG Leu	AAG Lys	TGT Cys	ATG Met	CTT Leu 615	GGG Gly	ATT Ile	CTC Leu	AAT Asn	GGA Gly 620	1934
TGC Cys	TGG Trp	ATT Ile	CTA Leu	AAA Lys 625	TTT Phe	GAA Glu	TGG Trp	GTA Val	AAA Lys 630	GCA Ala	TGT Cys	CTA Leu	CGA Arg	AGA Arg	AAA Lys	1982
GTA Val	TGT Cys	GAA Glu 640	CAG Gln	GAA Glu	GAA Glu	AAG Lys	TAT Tyr	GAA Glu 645	ATT Ile	CCT Pro	GAA Glu	GGT Gly	CCA Pro	CGC Arg	AGA Arg	2030
AGC Ser	AGG Arg	CTC Leu 655	AAC Asn	AGA Arg	GAA Glu	CAG Gln	CTG Leu 660	TTG Leu	CCA Pro	AAG Lys	CTG Leu	TTT Phe	GAT Asp	GGA Gly	TGC Cys	2078
TAC Tyr 670	TTC Phe	TAT Tyr	TTG Leu	TGG Trp	GGA Gly	ACC Thr 675	TTC Phe	AAA Lys	CAC His	CAT His	CCA Pro	AAG Lys	GAC Asp	AAC Asn	CTT Leu	2126
ATT Ile 685	AAG Lys	CTC Leu	GTC Val	ACT Thr	GCA Ala 690	GGT Gly	GGG Gly	GGC Gly	CAG Gln	ATC Ile 695	CTC Leu	AGT Ser	AGA Arg	AAG Lys	CCC Pro 700	2174

AAG CCA GAC AGT GAC GTG ACT CAG ACC ATC AAT ACA GTC GCA TAC CAT 2222  
 Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val Ala Tyr His  
 705 710 715  
 GCG AGA CCC GAT TCT GAT CAG CGC TTC TGC ACA CAG TAT ATC ATC TAT 2270  
 Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr Ile Ile Tyr  
 720 725 730  
 GAA GAT TTG TGT AAT TAT CAC CCA GAG AGG GTT CGG CAG GGC AAA GTC 2318  
 Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln Gly Lys Val  
 735 740 745  
 TGG AAG GCT CCT TCG AGC TGG TTT ATA GAC TGT GTG ATG TCC TTT GAG 2366  
 Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met Ser Phe Glu  
 750 755 760  
 TTG CTT CCT CTT GAC AGC TGAATATTAT ACCAGATGAA CATTTCAAAT 2414  
 Leu Leu Pro Leu Asp Ser  
 765 770  
 TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT TTTTAATGTT CACATTTTTA 2474  
 CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2510

## (2) INFORMATION FOR SEQ ID NO: 27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 770 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr  
 355 360 365  
 Ser Gly Arg Lys Asn Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser  
 370 375 380  
 Pro Gly Thr Pro Pro Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val  
 385 390 395 400  
 Met Ser Ser Pro Ser Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys  
 405 410 415  
 Arg Asn His Arg Gly Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly  
 420 425 430  
 Asp Ile Pro Ser Val Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn  
 435 440 445  
 Val Lys Asp His Ala Gly Trp Thr Pro Leu His Glu Ala Cys Asn His  
 450 455 460  
 Gly His Leu Lys Val Val Glu Leu Leu Leu Gln His Lys Ala Leu Val  
 465 470 475 480  
 Asn Thr Thr Gly Tyr Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys  
 485 490 495  
 Asn Gly His Val Asp Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser  
 500 505 510



Arg Asn Ala Val Asn Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp  
 515 520 525  
 Asp Glu Ser Met Lys Ser Leu Leu Leu Pro Glu Lys Asn Glu Ser  
 530 535 540  
 Ser Ser Ala Ser His Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp  
 545 550 555 560  
 Gly Pro Leu Val Leu Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys  
 565 570 575  
 Met Leu Ser Glu Leu Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu  
 580 585 590  
 Phe Asp Ser Thr Val Thr His Val Val Val Pro Gly Asp Ala Val Gln  
 595 600 605  
 Ser Thr Leu Lys Cys Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu  
 610 615 620  
 Lys Phe Glu Trp Val Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln  
 625 630 635 640  
 Glu Glu Lys Tyr Glu Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn  
 645 650 655  
 Arg Glu Gln Leu Leu Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu  
 660 665 670  
 Trp Gly Thr Phe Lys His His Pro Lys Asp Asn Leu Ile Lys Leu Val  
 675 680 685  
 Thr Ala Gly Gly Gly Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser  
 690 695 700  
 Asp Val Thr Gln Thr Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp  
 705 710 715 720  
 Ser Asp Gln Arg Phe Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys  
 725 730 735  
 Asn Tyr His Pro Glu Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro  
 740 745 750  
 Ser Ser Trp Phe Ile Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu  
 755 760 765  
 Asp Ser  
 770

## (2) INFORMATION FOR SEQ ID NO: 28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

(A) NAME/KEY: modified\_base\_  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

(A) NAME/KEY: modified\_base\_  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:28 and SEQ ID NO:29"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG  ATG  CCG  GAT  AAT  CGG  CAG  CCG  AGG  AAC  CGG  CAG  CCG      110
                  Met  Pro  Asp  Asn  Arg  Gln  Pro  Arg  Asn  Arg  Gln  Pro
                  1          5          10
AGG  ATC  CGC  TCC  GGG  AAC  GAG  CCT  CGT  TCC  GCG  CCC  GCC  ATG  GAA  CCG      158
Arg  Ile  Arg  Ser  Gly  Asn  Glu  Pro  Arg  Ser  Ala  Pro  Ala  Met  Glu  Pro
                  15          20          25
GAT  GGT  CGC  GGT  GCC  TGG  GCC  CAC  AGT  CGC  GCC  GCG  CTC  GAC  CGC  CTG      206
Asp  Gly  Arg  Gly  Ala  Trp  Ala  His  Ser  Arg  Ala  Ala  Leu  Asp  Arg  Leu
                  30          35          40
GAG  AAG  CTG  CTG  CGC  TGC  TCG  CGT  TGT  ACT  AAC  ATT  CTG  AGA  GAG  CCT      254
Glu  Lys  Leu  Leu  Arg  Cys  Ser  Arg  Cys  Thr  Asn  Ile  Leu  Arg  Glu  Pro
                  45          50          55          60
GTG  TGT  TTA  GGA  GGA  TGT  GAG  CAC  ATC  TTC  TGT  AGT  AAT  TGT  GTA  AGT      302
Val  Cys  Leu  Gly  Gly  Cys  Glu  His  Ile  Phe  Cys  Ser  Asn  Cys  Val  Ser
                  65          70          75
GAC  TGC  ATT  GGA  ACT  GGA  TGT  CCA  GTG  TGT  TAC  ACC  CCG  GCC  TGG  ATA      350
Asp  Cys  Ile  Gly  Thr  Gly  Cys  Pro  Val  Cys  Tyr  Thr  Pro  Ala  Trp  Ile
                  80          85          90
CAA  GAC  TTG  AAG  ATA  AAT  AGA  CAA  CTG  GAC  AGC  ATG  ATT  CAA  CTT  TGT      398
Gln  Asp  Leu  Lys  Ile  Asn  Arg  Gln  Leu  Asp  Ser  Met  Ile  Gln  Leu  Cys
                  95          100          105
AGT  AAG  CTT  CGA  AAT  TTG  CTA  CAT  GAC  AAT  GAG  CTG  TCA  GAT  TTG  AAA      446
Ser  Lys  Leu  Arg  Asn  Leu  Leu  His  Asp  Asn  Glu  Leu  Ser  Asp  Leu  Lys
                  110          115          120
GAA  GAT  AAA  CCT  AGG  AAA  AGT  TTG  TTT  AAT  GAT  GCA  GGA  AAC  AAG  AAG      494
Glu  Asp  Lys  Pro  Arg  Lys  Ser  Leu  Phe  Asn  Asp  Ala  Gly  Asn  Lys  Lys
                  125          130          135          140
AAT  TCA  ATT  AAA  ATG  TGG  TTT  AGC  CCT  CGA  AGT  AAG  RAA  GTC  AGA  TAT      542
Asn  Ser  Ile  Lys  Met  Trp  Phe  Ser  Pro  Arg  Ser  Lys  Xaa  Val  Arg  Tyr
                  145          150          155
GTT  GTG  AGT  AAA  GCT  TCA  GTG  CAA  ACC  CAG  CCT  GCA  ATA  AAA  AAA  GAT      590
Val  Val  Ser  Lys  Ala  Ser  Val  Gln  Thr  Gln  Pro  Ala  Ile  Lys  Lys  Asp
                  160          165          170
GCA  AGT  GCT  CAG  CAA  GAC  TCA  TAT  GAA  TTT  GTT  TCC  CCA  AGT  CCT  CCT      638
Ala  Ser  Ala  Gln  Asp  Ser  Tyr  Glu  Phe  Val  Ser  Pro  Ser  Pro  Pro
                  175          180          185

```

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC ATG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Met Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Met Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2531 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ix) FEATURE:
- (A) NAME/KEY: CDS
  - (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:30 and SEQ ID NO:31"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG  ATG  CCG  GAT  AAT  CGG  CAG  CCG  AGG  AAC  CGG  CAG  CCG      110
                  Met  Pro  Asp  Asn  Arg  Gln  Pro  Arg  Asn  Arg  Gln  Pro
                  1      5      10
AGG  ATC  CGC  TCC  GGG  AAC  GAG  CCT  CGT  TCC  GCG  CCC  GCC  ATG  GAA  CCG      158
Arg  Ile  Arg  Ser  Gly  Asn  Glu  Pro  Arg  Ser  Ala  Pro  Ala  Met  Glu  Pro
                  15      20      25
GAT  GGT  CGC  GGT  GCC  TGG  GCC  CAC  AGT  CGC  GCC  GCG  CTC  GAC  CGC  CTG      206
Asp  Gly  Arg  Gly  Ala  Trp  Ala  His  Ser  Arg  Ala  Ala  Leu  Asp  Arg  Leu
                  30      35      40
GAG  AAG  CTG  CTG  CGC  TGC  TCG  CGT  TGT  ACT  AAC  ATT  CTG  AGA  GAG  CCT      254
Glu  Lys  Leu  Leu  Arg  Cys  Ser  Arg  Cys  Thr  Asn  Ile  Leu  Arg  Glu  Pro
                  45      50      55      60
GTG  TGT  TTA  GGA  GGA  TGT  GAG  CAC  ATC  TTC  TGT  AGT  AAT  TGT  GTA  AGT      302
Val  Cys  Leu  Gly  Gly  Cys  Glu  His  Ile  Phe  Cys  Ser  Asn  Cys  Val  Ser
                  65      70      75
GAC  TGC  ATT  GGA  ACT  GGA  TGT  CCA  GTG  TGT  TAC  ACC  CCG  GCC  TGG  ATA      350
Asp  Cys  Ile  Gly  Thr  Gly  Cys  Pro  Val  Cys  Tyr  Thr  Pro  Ala  Trp  Ile
                  80      85      90
CAA  GAC  TTG  AAG  ATA  AAT  AGA  CAA  CTG  GAC  AGC  ATG  ATT  CAA  CTT  TGT      398
Gln  Asp  Leu  Lys  Ile  Asn  Arg  Gln  Leu  Asp  Ser  Met  Ile  Gln  Leu  Cys
                  95      100      105
AGT  AAG  CTT  CGA  AAT  TTG  CTA  CAT  GAC  AAT  GAG  CTG  TCA  GAT  TTG  AAA      446
Ser  Lys  Leu  Arg  Asn  Leu  Leu  His  Asp  Asn  Glu  Leu  Ser  Asp  Leu  Lys
                  110      115      120
GAA  GAT  AAA  CCT  AGG  AAA  AGT  TTG  TTT  AAT  GAT  GCA  GGA  AAC  AAG  AAG      494
Glu  Asp  Lys  Pro  Arg  Lys  Ser  Leu  Phe  Asn  Asp  Ala  Gly  Asn  Lys  Lys
                  125      130      135      140
AAT  TCA  ATT  AAA  ATG  TGG  TTT  AGC  CCT  CGA  AGT  AAG  RAA  GTC  AGA  TAT      542
Asn  Ser  Ile  Lys  Met  Trp  Phe  Ser  Pro  Arg  Ser  Lys  Xaa  Val  Arg  Tyr
                  145      150      155
GTT  GTG  AGT  AAA  GCT  TCA  GTG  CAA  ACC  CAG  CCT  GCA  ATA  AAA  AAA  GAT      590
Val  Val  Ser  Lys  Ala  Ser  Val  Gln  Thr  Gln  Pro  Ala  Ile  Lys  Lys  Asp
                  160      165      170
GCA  AGT  GCT  CAG  CAA  GAC  TCA  TAT  GAA  TTT  GTT  TCC  CCA  AGT  CCT  CCT      638
Ala  Ser  Ala  Gln  Gln  Asp  Ser  Tyr  Glu  Phe  Val  Ser  Pro  Ser  Pro  Pro
                  175      180      185

```



GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TCC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Ser Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTGCA CGGTTTGTGA GAGCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
           515                                  520                                  525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
           530                                  535                                  540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Ser Ser Val Met  
           545                                  550                                  555                                  560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
                                   565                                  570                                  575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
                                   580                                  585                                  590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
           595                                  600                                  605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
           610                                  615                                  620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
           625                                  630                                  635                                  640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
                                   645                                  650                                  655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
                                   660                                  665                                  670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
           675                                  680                                  685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
           690                                  695                                  700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
           705                                  710                                  715                                  720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
                                   725                                  730                                  735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
           740                                  745                                  750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
           755                                  760                                  765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
           770                                  775

## (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION: 531  
 (D) OTHER INFORMATION: /note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION: 153  
 (D) OTHER INFORMATION: /note= "Xaa = Glu or Lys for both  
 SEQ ID NO: 32 and SEQ ID NO: 33"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

```

CGAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG      110
      Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro
      1          5          10
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG      158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro
      15          20          25
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG      206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu
      30          35          40
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT      254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro
      45          50          55          60
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT      302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser
      65          70          75
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA      350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile
      80          85          90
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT      398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys
      95          100          105
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA      446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys
      110          115          120
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG      494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys
      125          130          135          140
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT      542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr
      145          150          155
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT      590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp
      160          165          170
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT      638
Ala Ser Ala Gln Gln Asp S r Tyr Glu Phe Val Ser Pro Ser Pro Pro
      175          180          185

```

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAC CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly His Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174



CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 33:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly His Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2531 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

(A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

(A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:34 and SEQ ID NO:35"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GCAGCTTCCC TGTGGTTTCC CGAGGCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT	60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG	110
Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro	
1 5 10	
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG	158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro	
15 20 25	
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG	206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu	
30 35 40	
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT	254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro	
45 50 55 60	
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT	302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser	
65 70 75	
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA	350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile	
80 85 90	
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT	398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys	
95 100 105	
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA	446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys	
110 115 120	
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG	494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys	
125 130 135 140	
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT	542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr	
145 150 155	
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT	590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp	
160 165 170	
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT	638
Ala Ser Ala Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro	
175 180 185	

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr L u Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA Glu 445	TAC Tyr	CTT Leu	TTA Leu	CAA Gln 450	AAT Asn 450	GGA Gly	AGT Ser	GAT Asp	CCA Pro	AAT Asn 455	GTT Val	AAA Lys	GAC Asp	CAT His	GCT Ala 460	1454
GGA Gly	TGG Trp	ACA Thr	CCA Pro	TTG Leu 465	CAT His	GAA Glu	GCT Ala	TGC Cys 470	AAT Asn 470	CAT His	GGG Gly	CAC His	CTG Leu	AAG Lys 475	GTA Val	1502
GTG Val	GAA Glu	TTA Leu	TTG Leu 480	CTC Leu	CAG Gln	CAT His	AAG Lys	GCA Ala 485	TTG Leu	GTG Val	AAC Asn	ACC Thr	ACC Thr	GGG Gly	TAT Tyr	1550
CAA Gln	AAT Asn	GAC Asp 495	TCA Ser	CCA Pro	CTT Leu	CAC His	GAT Asp 500	GCA Ala	GCC Ala	AAG Lys	AAT Asn	GGG Gly 505	CAC His	GTG Val	GAT Asp	1598
ATA Ile 510	GTC Val	AAG Lys	CTG Leu	TTA Leu	CTT Leu	TCC Ser 515	TAT Tyr	GGA Gly	GCC Ala	TCC Ser 520	AGA Arg	AAT Asn	GCT Ala	GTT Val	AAT Asn	1646
ATA Ile 525	TTT Phe	GGT Gly	CTG Leu	CGG Arg	CCT Pro 530	GTC Val	GAT Asp	TAT Tyr	ACA Thr	GAT Asp 535	GAT Asp	GAA Glu	AGT Ser	ATG Met	AAA Lys 540	1694
TCG Ser	CTA Leu	TTG Leu	CTG Leu	CTA Leu	CCA Pro	GAG Glu	AAG Lys	AAT Asn	GAA Glu 550	TCA Ser	TCC Ser	TCA Ser	GCT Ala	AGC Ser 555	CAC His	1742
TGC Cys	TCA Ser	GTA Val	ATG Met 560	AAC Asn	ACT Thr	GGG Gly	CAG Gln	CGT Arg 565	AGG Arg	GAT Asp	GGA Gly	CCT Pro	CTT Leu	GTA Val	CTT Leu	1790
ATA Ile	GGC Gly	AGT Ser 575	GGG Gly	CTG Leu	TCT Ser	TCA Ser	GAA Glu 580	CAA Gln	CAG Gln	AAA Lys	ATG Met	CTC Leu 585	AGT Ser	GAG Glu	CTT Leu	1838
GCA Ala	GTA Val	ATT Ile 590	CTT Leu	AAG Lys	GCT Ala	AAA Lys 595	AAA Lys	TAT Tyr	ACT Thr	GAG Glu	TTT Phe 600	GAC Asp	AGT Ser	ACA Thr	GTA Val	1886
ACT Thr 605	CAT His	GTT Val	GTT Val	GTT Val	CCT Pro 610	GGT Gly	GAT Asp	GCA Ala	GTT Val	CAA Gln 615	AGT Ser	ACC Thr	TTG Leu	AAG Lys	TGT Cys 620	1934
ATG Met	CTT Leu	GGG Gly	ATT Ile	CTC Leu	AAT Asn	GGA Gly	TGC Cys	TGG Trp	ATT Ile 630	CTA Leu	AAA Lys	TTT Phe	GAA Glu	TGG Trp 635	GTA Val	1982
AAA Lys	GCA Ala	TGT Cys	CTA Leu	CGA Arg	AGA Arg	AAA Lys	GTA Val	TGT Cys 645	GAA Glu	CAG Gln	GAA Glu	GAA Glu	AAG Lys 650	TAT Tyr	GAA Glu	2030
ATT Ile	CCT Pro	GAA Glu 655	GGT Gly	CCA Pro	CGC Arg	AGA Arg	AGC Ser 660	AGG Arg	CTC Leu	AAC Asn	AGA Arg	GAA Glu 665	CAG Gln	CTG Leu	TTG Leu	2078
CCA Pro	AAG Lys	CTG Leu	TTT Phe	GAT Asp	GGA Gly	TGC Cys	TAC Tyr	TTC Phe	TAT Tyr	TTG Leu	TGG Trp 680	GGA Gly	ACC Thr	TTC Phe	AAA Lys	2126
CAC His 685	CAT His	CCA Pro	AAG Lys	GAC Asp	AAC Asn	CTT Leu	ATT Ile	AAG Lys	CTC Leu	GTC Val 695	ACT Thr	GCA Ala	GGT Gly	GGG Gly	GGC Gly 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AAC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Asn Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTAA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510



Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Asn Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:36 and SEQ ID NO:37"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG  ATG  CCG  GAT  AAT  CGG  CAG  CCG  AGG  AAC  CGG  CAG  CCG      110
                  Met  Pro  Asp  Asn  Arg  Gln  Pro  Arg  Asn  Arg  Gln  Pro
                  1          5          10
AGG  ATC  CGC  TCC  GGG  AAC  GAG  CCT  CGT  TCC  GCG  CCC  GCC  ATG  GAA  CCG      158
Arg  Ile  Arg  Ser  Gly  Asn  Glu  Pro  Arg  Ser  Ala  Pro  Ala  Met  Glu  Pro
                  15          20          25
GAT  GGT  CGC  GGT  GCC  TGG  GCC  CAC  AGT  CGC  GCC  GCG  CTC  GAC  CGC  CTG      206
Asp  Gly  Arg  Gly  Ala  Trp  Ala  His  Ser  Arg  Ala  Ala  Leu  Asp  Arg  Leu
                  30          35          40
GAG  AAG  CTG  CTG  CGC  TGC  TCG  CGT  TGT  ACT  AAC  ATT  CTG  AGA  GAG  CCT      254
Glu  Lys  Leu  Leu  Arg  Cys  Ser  Arg  Cys  Thr  Asn  Ile  Leu  Arg  Glu  Pro
                  45          50          55          60
GTG  TGT  TTA  GGA  GGA  TGT  GAG  CAC  ATC  TTC  TGT  AGT  AAT  TGT  GTA  AGT      302
Val  Cys  Leu  Gly  Gly  Cys  Glu  His  Ile  Phe  Cys  Ser  Asn  Cys  Val  Ser
                  65          70          75
GAC  TGC  ATT  GGA  ACT  GGA  TGT  CCA  GTG  TGT  TAC  ACC  CCG  GCC  TGG  ATA      350
Asp  Cys  Ile  Gly  Thr  Gly  Cys  Pro  Val  Cys  Tyr  Thr  Pro  Ala  Trp  Ile
                  80          85          90
CAA  GAC  TTG  AAG  ATA  AAT  AGA  CAA  CTG  GAC  AGC  ATG  ATT  CAA  CTT  TGT      398
Gln  Asp  Leu  Lys  Ile  Asn  Arg  Gln  Leu  Asp  Ser  Met  Ile  Gln  Leu  Cys
                  95          100          105
AGT  AAG  CTT  CGA  AAT  TTG  CTA  CAT  GAC  AAT  GAG  CTG  TCA  GAT  TTG  AAA      446
Ser  Lys  Leu  Arg  Asn  Leu  Leu  His  Asp  Asn  Glu  Leu  Ser  Asp  Leu  Lys
                  110          115          120
GAA  GAT  AAA  CCT  AGG  AAA  AGT  TTG  TTT  AAT  GAT  GCA  GGA  AAC  AAG  AAG      494
Glu  Asp  Lys  Pro  Arg  Lys  Ser  Leu  Phe  Asn  Asp  Ala  Gly  Asn  Lys  Lys
                  125          130          135          140
AAT  TCA  ATT  AAA  ATG  TGG  TTT  AGC  CCT  CGA  AGT  AAG  RAA  GTC  AGA  TAT      542
Asn  Ser  Ile  Lys  Met  Trp  Phe  Ser  Pro  Arg  Ser  Lys  Xaa  Val  Arg  Tyr
                  145          150          155
GTT  GTG  AGT  AAA  GCT  TCA  GTG  CAA  ACC  CAG  CCT  GCA  ATA  AAA  AAA  GAT      590
Val  Val  Ser  Lys  Ala  Ser  Val  Gln  Thr  Gln  Pro  Ala  Ile  Lys  Lys  Asp
                  160          165          170
GCA  AGT  GCT  CAG  CAA  GAC  TCA  TAT  GAA  TTT  GTT  TCC  CCA  AGT  CCT  CCT      638
Ala  Ser  Ala  Gln  Gln  Asp  Ser  Tyr  Glu  Phe  Val  Ser  Pro  Ser  Pro  Pro
                  175          180          185

```

GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA TGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Cys Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AGC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTC CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Cys Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Ser Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2531 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 75..2405

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:531  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:153  
 (D) OTHER INFORMATION:/note= "Xaa = Glu or Lys for both  
 SEQ ID NO:38 and SEQ ID NO:39"

## (xi) SEQUENCE DESCRIPTION: SEQ ID-NO: 38:

```

GCAGCTTCCC TGTGGTTTCC CGAGGCCTCC TTGCTTCCCG CTCTCCGAGG AGCCTTTCAT      60
CCGAAGGCGG GACG ATG CCG GAT AAT CGG CAG CCG AGG AAC CGG CAG CCG      110
      Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro
      1          5          10
AGG ATC CGC TCC GGG AAC GAG CCT CGT TCC GCG CCC GCC ATG GAA CCG      158
Arg Ile Arg Ser Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro
      15          20          25
GAT GGT CGC GGT GCC TGG GCC CAC AGT CGC GCC GCG CTC GAC CGC CTG      206
Asp Gly Arg Gly Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu
      30          35          40
GAG AAG CTG CTG CGC TGC TCG CGT TGT ACT AAC ATT CTG AGA GAG CCT      254
Glu Lys Leu Leu Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro
      45          50          55          60
GTG TGT TTA GGA GGA TGT GAG CAC ATC TTC TGT AGT AAT TGT GTA AGT      302
Val Cys Leu Gly Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser
      65          70          75
GAC TGC ATT GGA ACT GGA TGT CCA GTG TGT TAC ACC CCG GCC TGG ATA      350
Asp Cys Ile Gly Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile
      80          85          90
CAA GAC TTG AAG ATA AAT AGA CAA CTG GAC AGC ATG ATT CAA CTT TGT      398
Gln Asp Leu Lys Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys
      95          100          105
AGT AAG CTT CGA AAT TTG CTA CAT GAC AAT GAG CTG TCA GAT TTG AAA      446
Ser Lys Leu Arg Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys
      110          115          120
GAA GAT AAA CCT AGG AAA AGT TTG TTT AAT GAT GCA GGA AAC AAG AAG      494
Glu Asp Lys Pro Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys
      125          130          135          140
AAT TCA ATT AAA ATG TGG TTT AGC CCT CGA AGT AAG RAA GTC AGA TAT      542
Asn Ser Ile Lys Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr
      145          150          155
GTT GTG AGT AAA GCT TCA GTG CAA ACC CAG CCT GCA ATA AAA AAA GAT      590
Val Val Ser Lys Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp
      160          165          170
GCA AGT GCT CAG CAA GAC TCA TAT GAA TTT GTT TCC CCA AGT CCT CCT      638
Ala Ser Ala Gln Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro
      175          180          185

```



GCA GAT GTT TCT GAG AGG GCT AAA AAG GCT TCT GCA AGA TCT GGA AAA Ala Asp Val Ser Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys 190 195 200	686
AAG CAA AAA AAG AAA ACT TTA GCT GAA ATC AAC CAA AAA TGG AAT TTA Lys Gln Lys Lys Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu 205 210 215 220	734
GAG GCA GAA AAA GAA GAT GGT GAA TTT GAC TCC AAA GAG GAA TCT AAG Glu Ala Glu Lys Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys 225 230 235	782
CAA AAG CTG GTA TCC TTC TGT AGC CAA CCA TCT GTT ATC TCC AGT CCT Gln Lys Leu Val Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro 240 245 250	830
CAG ATA AAT GGT GAA ATA GAC TTA CTA GCA AGT GGC TCC TTG ACA GAA Gln Ile Asn Gly Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu 255 260 265	878
TCT GAA TGT TTT GGA AGT TTA ACT GAA GTC TCT TTA CCA TTG GCT GAG Ser Glu Cys Phe Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu 270 275 280	926
CAA ATA GAG TCT CCA GAC ACT AAG AGC AGG AAT GAA GTA GTG ACT CCT Gln Ile Glu Ser Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro 285 290 295 300	974
GAG AAG GTC TGC AAA AAT TAT CTT ACA TCT AAG AAA TCT TTG CCA TTA Glu Lys Val Cys Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu 305 310 315	1022
GAA AAT AAT GGA AAA CGT GGC CAT CAC AAT AGA CTT TCC AGT CCC ATT Glu Asn Asn Gly Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile 320 325 330	1070
TCT AAG AGA TGT AGA ACC AGC ATT CTG AGC ACC AGT GGA GAT TTT GTT Ser Lys Arg Cys Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val 335 340 345	1118
AAG CAA ACC GTG CCC TCA GAA AAT ATA CCA TTG CCT GAA TGT TCT TCA Lys Gln Thr Val Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser 350 355 360	1166
CCA CCT TCA TGC AAA CGT AAA GTT GGT GGT ACA TCA GGG AGG AAA AAC Pro Pro Ser Cys Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn 365 370 375 380	1214
AGT AAC ATG TCC GAT GAA TTC ATT AGT CTT TCA CCA GGT ACA CCA CCT Ser Asn Met Ser Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro 385 390 395	1262
TCT ACA TTA AGT AGT TCA AGT TAC AGG CAA GTG ATG TCT AGT CCC TCA Ser Thr Leu Ser Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser 400 405 410	1310
GCA ATG AAG CTG TTG CCC AAT ATG GCT GTG AAA AGA AAT CAT AGA GGA Ala Met Lys Leu Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly 415 420 425	1358
GAG ACT TTG CTC CAT ATT GCT TCT ATT AAG GGC GAC ATA CCT TCT GTT Glu Thr Leu Leu His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val 430 435 440	1406

GAA TAC CTT TTA CAA AAT GGA AGT GAT CCA AAT GTT AAA GAC CAT GCT Glu Tyr Leu Leu Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala 445 450 455 460	1454
GGA TGG ACA CCA TTG CAT GAA GCT TGC AAT CAT GGG CAC CTG AAG GTA Gly Trp Thr Pro Leu His Glu Ala Cys Asn His Gly His Leu Lys Val 465 470 475	1502
GTG GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AAC ACC ACC GGG TAT Val Glu Leu Leu Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr 480 485 490	1550
CAA AAT GAC TCA CCA CTT CAC GAT GCA GCC AAG AAT GGG CAC GTG GAT Gln Asn Asp Ser Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp 495 500 505	1598
ATA GTC AAG CTG TTA CTT TCC TAT GGA GCC TCC AGA AAT GCT GTT AAT Ile Val Lys Leu Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn 510 515 520	1646
ATA TTT GGT CTG CGG CCT GTC GAT TAT ACA GAT GAT GAA AGT ATG AAA Ile Phe Gly Leu Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys 525 530 535 540	1694
TCG CTA TTG CTG CTA CCA GAG AAG AAT GAA TCA TCC TCA GCT AGC CAC Ser Leu Leu Leu Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His 545 550 555	1742
TGC TCA GTA ATG AAC ACT GGG CAG CGT AGG GAT GGA CCT CTT GTA CTT Cys Ser Val Met Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu 560 565 570	1790
ATA GGC AGT GGG CTG TCT TCA GAA CAA CAG AAA ATG CTC AGT GAG CTT Ile Gly Ser Gly Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu 575 580 585	1838
GCA GTA ATT CTT AAG GCT AAA AAA TAT ACT GAG TTT GAC AGT ACA GTA Ala Val Ile Leu Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val 590 595 600	1886
ACT CAT GTT GTT GTT CCT GGT GAT GCA GTT CAA AGT ACC TTG AAG TGT Thr His Val Val Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys 605 610 615 620	1934
ATG CTT GGG ATT CTC AAT GGA TGC TGG ATT CTA AAA TTT GAA TGG GTA Met Leu Gly Ile Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val 625 630 635	1982
AAA GCA TGT CTA CGA AGA AAA GTA TGT GAA CAG GAA GAA AAG TAT GAA Lys Ala Cys Leu Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu 640 645 650	2030
ATT CCT GAA GGT CCA CGC AGA AGC AGG CTC AAC AGA GAA CAG CTG TTG Ile Pro Glu Gly Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu 655 660 665	2078
CCA AAG CTG TTT GAT GGA TGC TAC TTC TAT TTG TGG GGA ACC TTC AAA Pro Lys Leu Phe Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys 670 675 680	2126
CAC CAT CCA AAG GAC AAC CTT ATT AAG CTC GTC ACT GCA GGT GGG GGC His His Pro Lys Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly 685 690 695 700	2174

CAG ATC CTC AGT AGA AAG CCC AAG CCA GAC AGT GAC GTG ACT CAG ACC 2222  
 Gln Ile Leu Ser Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr  
 705 710 715  
 ATC AAT ACA GTC GCA TAC CAT GCG AGA CCC GAT TCT GAT CAG CGC TTC 2270  
 Ile Asn Thr Val Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe  
 720 725 730  
 TGC ACA CAG TAT ATC ATC TAT GAA GAT TTG TGT AAT TAT CAC CCA GAG 2318  
 Cys Thr Gln Tyr Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu  
 735 740 745  
 AGG GTT CGG CAG GGC AAA GTC TGG AAG GCT CCT TCG AAC TGG TTT ATA 2366  
 Arg Val Arg Gln Gly Lys Val Trp Lys Ala Pro Ser Asn Trp Phe Ile  
 750 755 760  
 GAC TGT GTG ATG TCC TTT GAG TTG CTT CCT CTT GAC AGC TGAATATTAT 2415  
 Asp Cys Val Met Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 765 770 775  
 ACCAGATGAA CATTTCAAAT TGAATTTGCA CGGTTTGTGA GAGCCCAGTC ATTGTACTGT 2475  
 TTTTAATGTT CACATTTTTA CAAATAGGTA GAGTCATTCA TATTTGTCTT TGAATC 2531

## (2) INFORMATION FOR SEQ ID NO: 39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

Met Pro Asp Asn Arg Gln Pro Arg Asn Arg Gln Pro Arg Ile Arg Ser  
 1 5 10 15  
 Gly Asn Glu Pro Arg Ser Ala Pro Ala Met Glu Pro Asp Gly Arg Gly  
 20 25 30  
 Ala Trp Ala His Ser Arg Ala Ala Leu Asp Arg Leu Glu Lys Leu Leu  
 35 40 45  
 Arg Cys Ser Arg Cys Thr Asn Ile Leu Arg Glu Pro Val Cys Leu Gly  
 50 55 60  
 Gly Cys Glu His Ile Phe Cys Ser Asn Cys Val Ser Asp Cys Ile Gly  
 65 70 75 80  
 Thr Gly Cys Pro Val Cys Tyr Thr Pro Ala Trp Ile Gln Asp Leu Lys  
 85 90 95  
 Ile Asn Arg Gln Leu Asp Ser Met Ile Gln Leu Cys Ser Lys Leu Arg  
 100 105 110  
 Asn Leu Leu His Asp Asn Glu Leu Ser Asp Leu Lys Glu Asp Lys Pro  
 115 120 125  
 Arg Lys Ser Leu Phe Asn Asp Ala Gly Asn Lys Lys Asn Ser Ile Lys  
 130 135 140  
 Met Trp Phe Ser Pro Arg Ser Lys Xaa Val Arg Tyr Val Val Ser Lys  
 145 150 155 160

Ala Ser Val Gln Thr Gln Pro Ala Ile Lys Lys Asp Ala Ser Ala Gln  
 165 170 175  
 Gln Asp Ser Tyr Glu Phe Val Ser Pro Ser Pro Pro Ala Asp Val Ser  
 180 185 190  
 Glu Arg Ala Lys Lys Ala Ser Ala Arg Ser Gly Lys Lys Gln Lys Lys  
 195 200 205  
 Lys Thr Leu Ala Glu Ile Asn Gln Lys Trp Asn Leu Glu Ala Glu Lys  
 210 215 220  
 Glu Asp Gly Glu Phe Asp Ser Lys Glu Glu Ser Lys Gln Lys Leu Val  
 225 230 235 240  
 Ser Phe Cys Ser Gln Pro Ser Val Ile Ser Ser Pro Gln Ile Asn Gly  
 245 250 255  
 Glu Ile Asp Leu Leu Ala Ser Gly Ser Leu Thr Glu Ser Glu Cys Phe  
 260 265 270  
 Gly Ser Leu Thr Glu Val Ser Leu Pro Leu Ala Glu Gln Ile Glu Ser  
 275 280 285  
 Pro Asp Thr Lys Ser Arg Asn Glu Val Val Thr Pro Glu Lys Val Cys  
 290 295 300  
 Lys Asn Tyr Leu Thr Ser Lys Lys Ser Leu Pro Leu Glu Asn Asn Gly  
 305 310 315 320  
 Lys Arg Gly His His Asn Arg Leu Ser Ser Pro Ile Ser Lys Arg Cys  
 325 330 335  
 Arg Thr Ser Ile Leu Ser Thr Ser Gly Asp Phe Val Lys Gln Thr Val  
 340 345 350  
 Pro Ser Glu Asn Ile Pro Leu Pro Glu Cys Ser Ser Pro Pro Ser Cys  
 355 360 365  
 Lys Arg Lys Val Gly Gly Thr Ser Gly Arg Lys Asn Ser Asn Met Ser  
 370 375 380  
 Asp Glu Phe Ile Ser Leu Ser Pro Gly Thr Pro Pro Ser Thr Leu Ser  
 385 390 395 400  
 Ser Ser Ser Tyr Arg Gln Val Met Ser Ser Pro Ser Ala Met Lys Leu  
 405 410 415  
 Leu Pro Asn Met Ala Val Lys Arg Asn His Arg Gly Glu Thr Leu Leu  
 420 425 430  
 His Ile Ala Ser Ile Lys Gly Asp Ile Pro Ser Val Glu Tyr Leu Leu  
 435 440 445  
 Gln Asn Gly Ser Asp Pro Asn Val Lys Asp His Ala Gly Trp Thr Pro  
 450 455 460  
 Leu His Glu Ala Cys Asn His Gly His Leu Lys Val Val Glu Leu Leu  
 465 470 475 480  
 Leu Gln His Lys Ala Leu Val Asn Thr Thr Gly Tyr Gln Asn Asp Ser  
 485 490 495  
 Pro Leu His Asp Ala Ala Lys Asn Gly His Val Asp Ile Val Lys Leu  
 500 505 510

Leu Leu Ser Tyr Gly Ala Ser Arg Asn Ala Val Asn Ile Phe Gly Leu  
 515 520 525  
 Arg Pro Val Asp Tyr Thr Asp Asp Glu Ser Met Lys Ser Leu Leu Leu  
 530 535 540  
 Leu Pro Glu Lys Asn Glu Ser Ser Ser Ala Ser His Cys Ser Val Met  
 545 550 555 560  
 Asn Thr Gly Gln Arg Arg Asp Gly Pro Leu Val Leu Ile Gly Ser Gly  
 565 570 575  
 Leu Ser Ser Glu Gln Gln Lys Met Leu Ser Glu Leu Ala Val Ile Leu  
 580 585 590  
 Lys Ala Lys Lys Tyr Thr Glu Phe Asp Ser Thr Val Thr His Val Val  
 595 600 605  
 Val Pro Gly Asp Ala Val Gln Ser Thr Leu Lys Cys Met Leu Gly Ile  
 610 615 620  
 Leu Asn Gly Cys Trp Ile Leu Lys Phe Glu Trp Val Lys Ala Cys Leu  
 625 630 635 640  
 Arg Arg Lys Val Cys Glu Gln Glu Glu Lys Tyr Glu Ile Pro Glu Gly  
 645 650 655  
 Pro Arg Arg Ser Arg Leu Asn Arg Glu Gln Leu Leu Pro Lys Leu Phe  
 660 665 670  
 Asp Gly Cys Tyr Phe Tyr Leu Trp Gly Thr Phe Lys His His Pro Lys  
 675 680 685  
 Asp Asn Leu Ile Lys Leu Val Thr Ala Gly Gly Gly Gln Ile Leu Ser  
 690 695 700  
 Arg Lys Pro Lys Pro Asp Ser Asp Val Thr Gln Thr Ile Asn Thr Val  
 705 710 715 720  
 Ala Tyr His Ala Arg Pro Asp Ser Asp Gln Arg Phe Cys Thr Gln Tyr  
 725 730 735  
 Ile Ile Tyr Glu Asp Leu Cys Asn Tyr His Pro Glu Arg Val Arg Gln  
 740 745 750  
 Gly Lys Val Trp Lys Ala Pro Ser Asn Trp Phe Ile Asp Cys Val Met  
 755 760 765  
 Ser Phe Glu Leu Leu Pro Leu Asp Ser  
 770 775

## (2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1083 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 37..819

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:346..378  
 (D) OTHER INFORMATION:/note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:690  
 (D) OTHER INFORMATION:/note= "W = A or T"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:104  
 (D) OTHER INFORMATION:/note= "Xaa = Ala or Ser or Pro or Thr for both SEQ ID NO:40 and 41"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:114  
 (D) OTHER INFORMATION:/note= "Xaa = Gly for both SEQ ID NO:40 and SEQ ID NO:41"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:218  
 (D) OTHER INFORMATION:/note= "Xaa = Ala for both SEQ ID NO:40 and SEQ ID NO:41"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CAGTTGCAGG CAGACGGAGC AGAGCGGTCA GGGATC ATG AGG GAG AGT GCG TTG	54
Met Arg Glu Ser Ala Leu	
1 5	
GAG CCG GGG CCT GTG CCC GAG GCG CCG GCG GGG GGT CCC GTG CAC GCC	102
Glu Pro Gly Pro Val Pro Glu Ala Pro Ala Gly Gly Pro Val His Ala	
10 15 20	
GTG ACG GTG GTG ACC CTG CTG GAG AAG CTG GCC TCC ATG CTG GAG ACT	150
Val Thr Val Val Thr Leu Leu Glu Lys Leu Ala Ser Met Leu Glu Thr	
25 30 35	
CTG CGG GAG CGG CAG GGA GGC CTG GCT CGA AGG CAG GGA GGC CTG GCA	198
Leu Arg Glu Arg Gln Gly Leu Ala Arg Arg Gln Gly Gly Leu Ala	
40 45 50	
GGG TCC GTG CGC CGC ATC CAG AGC GGC CTG GGC GCT CTG AGT CGC AGC	246
Gly Ser Val Arg Arg Ile Gln Ser Gly Leu Gly Ala Leu Ser Arg Ser	
55 60 65 70	
CAC GAC ACC ACC AGC AAC ACC TTG GCG CAG CTG CTG GCC AAG GCG GAG	294
His Asp Thr Thr Ser Asn Thr Leu Ala Gln Leu Leu Ala Lys Ala Glu	
75 80 85	
CGC GTG AGC TCG CAC GCC AAC GCC GCC CAA GAG CGC GCG GTG CGC CGC	342
Arg Val Ser Ser His Ala Asn Ala Ala Gln Glu Arg Ala Val Arg Arg	
90 95 100	
GCA RCC CAG GTG CAG CGG CTG GAG GCC AAC CAC GGR CTG CTG GTG GCG	390
Ala Xaa Gln Val Gln Arg Leu Glu Ala Asn His Xaa Leu Leu Val Ala	
105 110 115	
CGC GGG AAG CTC CAC GTT CTG CTC TTC AAG GAG GAG GGT GAA GTC CCA	438
Arg Gly Lys Leu His Val Leu Leu Phe Lys Glu Glu Gly Glu Val Pro	
120 125 130	

GCC AGC GCT TTC CAG AAG GCA CCA GAG CCC TTG GGC CCG GCG GAC CAG 486  
 Ala Ser Ala Phe Gln Lys Ala Pro Glu Pro Leu Gly Pro Ala Asp Gln 150  
 135 140 145  
 TCC GAG CTG GGC CCA GAG CAG CTG GAG GCC GAA GTT GGA GAG AGC TCG 534  
 Ser Glu Leu Gly Pro Glu Gln Leu Glu Ala Glu Val Gly Glu Ser Ser 165  
 155 160  
 GAC GAG GAG CCG GTG GAG TCC AGG GCC CAG CGG CTG CGG CGC ACC GGA 582  
 Asp Glu Glu Pro Val Glu Ser Arg Ala Gln Arg Leu Arg Arg Thr Gly 180  
 170 175  
 TTG CAG AAG GTA CAG AGC CTC CGA AGG GCC CTT TCG GGC CGG AAA GGC 630  
 Leu Gln Lys Val Gln Ser Leu Arg Arg Ala Leu Ser Gly Arg Lys Gly 195  
 185 190  
 CCT GCA GCG CCA CCG CCC ACC CCG GTC AAG CCG CCT CGC CTT GGG CCT 678  
 Pro Ala Ala Pro Pro Pro Thr Pro Val Lys Pro Pro Arg Leu Gly Pro 210  
 200 205  
 GGC CGG AGC GCW GAA GCC CAG CCG GAA GCC CAG CCT GCG CTG GAG CCC 726  
 Gly Arg Ser Xaa Glu Ala Gln Pro Glu Ala Gln Pro Ala Leu Glu Pro 230  
 215 220 225  
 ACG CTG GAG CCA GAG CCT CCG CAG GAC ACC GAG GAA GAT CCC GGG AGA 774  
 Thr Leu Glu Pro Glu Pro Pro Gln Asp Thr Glu Glu Asp Pro Gly Arg 245  
 235 240  
 CCT GGG GCT GCC GAA GAA GCT CTG CTC CAA ATG GAG AGT GTA GCC 819  
 Pro Gly Ala Ala Glu Glu Ala Leu Leu Gln Met Glu Ser Val Ala 260  
 250 255  
 TGAGGGCTGG TGTTCCTGC CTCCCCTGTG CTTGTGCCTT GTCCCAAAT AAATCCTTTC 879  
 AGAATGTAGC ACTCAGCCCC TAATAAGGAG CGAATCCTAC ATCCACCAAG GCGGGCGCTC 939  
 TGGCCCTCCC TTCCTTAAGC CCAGTCCTGT GTCCTCTGAA AGAGGTGCAG TCACTCACAC 999  
 CTGCTTGCGC TCACCATCAA TAAAAGTAAT TTCACCCGAA AAAAAAAAAA AAAAAAAAAA 1059  
 AAAAAAAAAA AAAAAAAAAA AAAA 1083

## (2) INFORMATION FOR SEQ ID NO: 41:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 261 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

Met Arg Glu Ser Ala Leu Glu Pro Gly Pro Val Pro Glu Ala Pro Ala  
 1 5 10 15  
 Gly Gly Pro Val His Ala Val Thr Val Val Thr Leu Leu Glu Lys Leu  
 20 25 30  
 Ala Ser Met Leu Glu Thr Leu Arg Glu Arg Gln Gly Gly Leu Ala Arg  
 35 40 45  
 Arg Gln Gly Gly Leu Ala Gly Ser Val Arg Arg Ile Gln Ser Gly Leu  
 50 55 60

Gly Ala Leu Ser Arg Ser His Asp Thr Thr Ser Asn Thr Leu Ala Gln  
 65 70 75 80  
 Leu Leu Ala Lys Ala Glu Arg Val Ser Ser His Ala Asn Ala Ala Gln  
 85 90 95  
 Glu Arg Ala Val Arg Arg Ala Xaa Gln Val Gln Arg Leu Glu Ala Asn  
 100 105 110  
 His Xaa Leu Leu Val Ala Arg Gly Lys Leu His Val Leu Leu Phe Lys  
 115 120 125  
 Glu Glu Gly Glu Val Pro Ala Ser Ala Phe Gln Lys Ala Pro Glu Pro  
 130 135 140  
 Leu Gly Pro Ala Asp Gln Ser Glu Leu Gly Pro Glu Gln Leu Glu Ala  
 145 150 155 160  
 Glu Val Gly Glu Ser Ser Asp Glu Glu Pro Val Glu Ser Arg Ala Gln  
 165 170 175  
 Arg Leu Arg Arg Thr Gly Leu Gln Lys Val Gln Ser Leu Arg Arg Ala  
 180 185 190  
 Leu Ser Gly Arg Lys Gly Pro Ala Ala Pro Pro Pro Thr Pro Val Lys  
 195 200 205  
 Pro Pro Arg Leu Gly Pro Gly Arg Ser Xaa Glu Ala Gln Pro Glu Ala  
 210 215 220  
 Gln Pro Ala Leu Glu Pro Thr Leu Glu Pro Glu Pro Pro Gln Asp Thr  
 225 230 235 240  
 Glu Glu Asp Pro Gly Arg Pro Gly Ala Ala Glu Glu Ala Leu Leu Gln  
 245 250 255  
 Met Glu Ser Val Ala  
 260

## (2) INFORMATION FOR SEQ ID NO: 42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1326 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..666

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGA ATT CCT GCT GTA CCA TGC CAT GCT CCC TCT CAT TCT GAA TCT CAG	48
Gly Ile Pro Ala Val Pro Cys His Ala Pro Ser His Ser Glu Ser Gln	
1 5 10 15	
GCA ACT CCT CAT TCT AGT TAT GGC TTA TGT ACC TCC ACC CCA GTC TGG	96
Ala Thr Pro His Ser Ser Tyr Gly Leu Cys Thr Ser Thr Pro Val Trp	
20 25 30	
TCA CTT CAG CGG CCA CCC TGC CCT CCA AAG GTT CAT TCT GAA GTT CAA	144
Ser Leu Gln Arg Pro Pro Cys Pro Pro Lys Val His Ser Glu Val Gln	
35 40 45	



ACT GAT GGC AAC AGT CAG TTT GCA TCA CAA GGT AAA ACA GTT TCT GCA Thr Asp Gly Asn Ser Gln Phe Ala Ser Gln Gly Lys Thr Val Ser Ala 50 55 60	192
ACC TGT ACT GAT GTT CTA CGG AAT TCA TTT AAT ACC AGT CCT GGA GTT Thr Cys Thr Asp Val Leu Arg Asn Ser Phe Asn Thr Ser Pro Gly Val 65 70 75 80	240
CCA TGT AGC CTG CCC AAA ACT GAC ATA TCA GCT ATT CCA ACA TTG CAG Pro Cys Ser Leu Pro Lys Thr Asp Ile Ser Ala Ile Pro Thr Leu Gln 85 90 95	288
CAA CTG GGC CTT GTT AAT GGA ATT CTG CCA CAA CAA GGA ATT CAT AAG Gln Leu Gly Leu Val Asn Gly Ile Leu Pro Gln Gln Gly Ile His Lys 100 105 110	336
GAA ACA GAC CTA CTA AAA TGT ATT CAA ACA TAT TTG TCT CTT TTT CGA Glu Thr Asp Leu Leu Lys Cys Ile Gln Thr Tyr Leu Ser Leu Phe Arg 115 120 125	384
TCT CAT GGA AAA GAA ACG CAT CTG GAC AGT CAG ACA CAC CGA AGC CCT Ser His Gly Lys Glu Thr His Leu Asp Ser Gln Thr His Arg Ser Pro 130 135 140	432
ACT CAG TCA CAA CCA GCT TTC TTG GCC ACT AAT GAA GAA AAA TGT GCC Thr Gln Ser Gln Pro Ala Phe Leu Ala Thr Asn Glu Glu Lys Cys Ala 145 150 155 160	480
AGA GAG CAA ATT AGA GAG GCC ACA AGT GAA AGA AAG GAT TTA AAC ATA Arg Glu Gln Ile Arg Glu Ala Thr Ser Glu Arg Lys Asp Leu Asn Ile 165 170 175	528
CAT GTG CGA GAT ACA AAA ACA GTG AAG GAT GTA CAG AAG GCA AAA AAT His Val Arg Asp Thr Lys Thr Val Lys Asp Val Gln Lys Ala Lys Asn 180 185 190	576
GTG AAC AAG ACA GCT GAA AAA GTT AGA ATT ATA AAA TAT TTG TTG GGA Val Asn Lys Thr Ala Glu Lys Val Arg Ile Ile Lys Tyr Leu Leu Gly 195 200 205	624
GAG CTC AAG GCC CTG GTA GCA GAA CAA GGT AGA TGG GAC TTA Glu Leu Lys Ala Leu Val Ala Glu Gln Gly Arg Trp Asp Leu 210 215 220	666
TAAC TTTCTG TAGTATGGTG TTATACTAAA TAGCAATGTC ATGTTATTTA GCTATCATTT	726
AAATGGAGTT TGTGGTATTT TCCATAGAAC TGTGTTTTGA GCTAATAAGA AAATGAGTTC	786
TACTTATTGT ATTATTTTTT AAGTTTTGAT CCCTTCTTTC CTGTGGATTT AAAATGCGTT	846
TGAGAATATC AAACATTCAG TCTTTTGCTT GCAAGTGTGT ATTTATTCTG CTTGATAATA	906
GACCTTGAAA AGAGTCAACC AAAGAGAATT TGGACAGATA AAAATTTTAA TTAGAGAATG	966
CCTATAAATG ATTAAC TCCC TGAGTAGACT GATTATTCTT CCTGTTTTAA AAAGATGCAG	1026
AGAATTCTTT CCTGTCACTT CTTTAATAGC CAACTGTTAG ATTGTTTAAC AAATCTCACT	1086
TTGAGAAGTA ACGCATACCT TCTTATGCCC TTTTCAGTGT ATTTTITAGGA CTTTTTTTCT	1146
TAAATCAAGG TGTTTCTGAG CCAGATTCTA TTCATTTGTT TCCATTCTGT ATATGTATTC	1206
TATAGTAATG GCTTTTGCTT GAAATGAGTT ACAGTTTTGT CATCTTGGAA ACACAGTAAT	1266
TGATTTTGGA AGCATTGATT GAATACCTAA CGTTTGCAGA CCAAAAAAAAA AAAAAAAAAA	1326

## (2) INFORMATION FOR SEQ ID NO: 43:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID-NO: 43:

```

Gly Ile Pro Ala Val Pro Cys His Ala Pro Ser His Ser Glu Ser Gln
 1             5             10             15
Ala Thr Pro His Ser Ser Tyr Gly Leu Cys Thr Ser Thr Pro Val Trp
          20             25             30
Ser Leu Gln Arg Pro Pro Cys Pro Pro Lys Val His Ser Glu Val Gln
          35             40             45
Thr Asp Gly Asn Ser Gln Phe Ala Ser Gln Gly Lys Thr Val Ser Ala
 50             55             60
Thr Cys Thr Asp Val Leu Arg Asn Ser Phe Asn Thr Ser Pro Gly Val
 65             70             75             80
Pro Cys Ser Leu Pro Lys Thr Asp Ile Ser Ala Ile Pro Thr Leu Gln
          85             90             95
Gln Leu Gly Leu Val Asn Gly Ile Leu Pro Gln Gln Gly Ile His Lys
          100             105             110
Glu Thr Asp Leu Leu Lys Cys Ile Gln Thr Tyr Leu Ser Leu Phe Arg
          115             120             125
Ser His Gly Lys Glu Thr His Leu Asp Ser Gln Thr His Arg Ser Pro
          130             135             140
Thr Gln Ser Gln Pro Ala Phe Leu Ala Thr Asn Glu Glu Lys Cys Ala
          145             150             155             160
Arg Glu Gln Ile Arg Glu Ala Thr Ser Glu Arg Lys Asp Leu Asn Ile
          165             170             175
His Val Arg Asp Thr Lys Thr Val Lys Asp Val Gln Lys Ala Lys Asn
          180             185             190
Val Asn Lys Thr Ala Glu Lys Val Arg Ile Ile Lys Tyr Leu Leu Gly
          195             200             205
Glu Leu Lys Ala Leu Val Ala Glu Gln Gly Arg Trp Asp Leu
          210             215             220

```

## (2) INFORMATION FOR SEQ ID NO: 44:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 834 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION:1..693

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GAA AAT GAA AAA ATA GTG GAA ACA TAC AGG GGA AAG GAA ACA GAA TAT	48
Glu Asn Glu Lys Ile Val Glu Thr Tyr Arg Gly Lys Glu Thr Glu Tyr	
1 5 10 15	
CAA GCG TTA CAA GAG ACT AAC ATG AAG TTT TCT ATG ATG CTG CGA GAA	96
Gln Ala Leu Gln Glu Thr Asn Met Lys Phe Ser Met Met Leu Arg Glu	
20 25 30	
AAA GAG TTT GAG TGC CAC TCA ATG AAG GAG AAG GCT CTT GCT TTT GAA	144
Lys Glu Phe Glu Cys His Ser Met Lys Glu Lys Ala Leu Ala Phe Glu	
35 40 45	
CAG CTA TTG AAA GAG AAA GAA CAG GGC AAG ACT GGA GAG TTA AAT CAG	192
Gln Leu Leu Lys Glu Lys Glu Gln Gly Lys Thr Gly Glu Leu Asn Gln	
50 55 60	
CTT TTA AAT GCA GTT AAA TCA ATG CAG GAG AAG ACA GTT GTG TTT CAA	240
Leu Leu Asn Ala Val Lys Ser Met Gln Glu Lys Thr Val Val Phe Gln	
65 70 75 80	
CAG GAG AGA GAC CAA GTC ATG TTG GCC CTG AAA CAA AAA CAA ATG GAA	288
Gln Glu Arg Asp Gln Val Met Leu Ala Leu Lys Gln Lys Gln Met Glu	
85 90 95	
AAT ACT GCC CTA CAG AAT GAG GTT CAA CGT TTA CGT GAC AAA GAA TTT	336
Asn Thr Ala Leu Gln Asn Glu Val Gln Arg Leu Arg Asp Lys Glu Phe	
100 105 110	
CGT TCA AAC CAA GAG CTA GAG AGA TTG CGT AAT CAT CTT TTA GAA TCA	384
Arg Ser Asn Gln Glu Leu Glu Arg Leu Arg Asn His Leu Leu Glu Ser	
115 120 125	
GAA GAT TCT TAT ACC CGT GAA GCT TTG GCT GCA GAA GAT AGA GAG GCT	432
Glu Asp Ser Tyr Thr Arg Glu Ala Leu Ala Ala Glu Asp Arg Glu Ala	
130 135 140	
AAA CTA AGA AAG AAA GTC ACA GTA TTG GAG GAA AAG CTA GTT TCA TCC	480
Lys Leu Arg Lys Lys Val Thr Val Leu Glu Glu Lys Leu Val Ser Ser	
145 150 155 160	
TCT AAT GCA ATG GAA AAT GCA AGC CAT CAA GCC AGT GTG CAG GTA GAG	528
Ser Asn Ala Met Glu Asn Ala Ser His Gln Ala Ser Val Gln Val Glu	
165 170 175	
TCA TTG CAA GAA CAG TTG AAT GTA GTT TCC AAG CAA AGG GAT GAA ACT	576
Ser Leu Gln Glu Gln Leu Asn Val Val Ser Lys Gln Arg Asp Glu Thr	
180 185 190	
GCG CTG CAG CTT TCT GTC TCT CAG GAA CAA GTA AAG CAG TAT GCT CTG	624
Ala Leu Gln Leu Ser Val Ser Gln Glu Gln Val Lys Gln Tyr Ala Leu	
195 200 205	
TCA CTG GCC AAC CTG CAG ATG GTA CTA GAG CAT TTC CAA CAA GAG GAA	672
Ser Leu Ala Asn Leu Gln Met Val Leu Glu His Phe Gln Gln Glu Glu	
210 215 220	
AAA GCT ATG TAT TCT GCT GAA CTCGAAAAGC AAAAAAAAAA AAAAAAACT	723
Lys Ala Met Tyr Ser Ala Glu	
225 230	
CGAGAGATCT ATGAATCGTA GATACTGAAA AACCCCGCAA GTTCACTTCA ACTGTGCATC	783
GTGCACCATC TCAATTTCTT TCATTTATAC ATCGTTTTC CTTCTTTTAT G	834

## (2) INFORMATION FOR SEQ ID NO: 45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```

Glu Asn Glu Lys Ile Val Glu Thr Tyr Arg Gly Lys Glu Thr Glu Tyr
 1             5             10             15
Gln Ala Leu Gln Glu Thr Asn Met Lys Phe Ser Met Met Leu Arg Glu
      20             25             30
Lys Glu Phe Glu Cys His Ser Met Lys Glu Lys Ala Leu Ala Phe Glu
      35             40             45
Gln Leu Leu Lys Glu Lys Glu Gln Gly Lys Thr Gly Glu Leu Asn Gln
      50             55             60
Leu Leu Asn Ala Val Lys Ser Met Gln Glu Lys Thr Val Val Phe Gln
      65             70             75             80
Gln Glu Arg Asp Gln Val Met Leu Ala Leu Lys Gln Lys Gln Met Glu
      85             90             95
Asn Thr Ala Leu Gln Asn Glu Val Gln Arg Leu Arg Asp Lys Glu Phe
      100            105            110
Arg Ser Asn Gln Glu Leu Glu Arg Leu Arg Asn His Leu Leu Glu Ser
      115            120            125
Glu Asp Ser Tyr Thr Arg Glu Ala Leu Ala Ala Glu Asp Arg Glu Ala
      130            135            140
Lys Leu Arg Lys Lys Val Thr Val Leu Glu Glu Lys Leu Val Ser Ser
      145            150            155            160
Ser Asn Ala Met Glu Asn Ala Ser His Gln Ala Ser Val Gln Val Glu
      165            170            175
Ser Leu Gln Glu Gln Leu Asn Val Val Ser Lys Gln Arg Asp Glu Thr
      180            185            190
Ala Leu Gln Leu Ser Val Ser Gln Glu Gln Val Lys Gln Tyr Ala Leu
      195            200            205
Ser Leu Ala Asn Leu Gln Met Val Leu Glu His Phe Gln Gln Glu Glu
      210            215            220
Lys Ala Met Tyr Ser Ala Glu
      225            230

```

## (2) INFORMATION FOR SEQ ID NO: 46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 898 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..816

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CTC GTG CCG CGA GAC CCT GAG CCA GAG CAG GCT GGG CCC AGC TCT GGA	48
Leu Val Pro Arg Asp Pro Glu Pro Glu Gln Ala Gly Pro Ser Ser Gly	
1 5 10 15	
GTC ACG AAC AGG TGC CCG TTC CTC CTG GAC AAT TGC CTT GGC ACA TCT	96
Val Thr Asn Arg Cys Pro Phe Leu Leu Asp Asn Cys Leu Gly Thr Ser	
20 25 30	
CAG TGG CCC CCA AGG CGA CGA CGC AAG CAG CTG TTC ACC CTG CAG ACG	144
Gln Trp Pro Pro Arg Arg Arg Arg Lys Gln Leu Phe Thr Leu Gln Thr	
35 40 45	
GTG AAC TCC AAT GGG ACC AGC GAC CGC ACA ACC TCC CCT GAA GAA GTC	192
Val Asn Ser Asn Gly Thr Ser Asp Arg Thr Thr Ser Pro Glu Glu Val	
50 55 60	
CAT GCC CAG CCG TAC ATT GCT ATC GAC TGG GAG CCA GAG ATG AAG AAG	240
His Ala Gln Pro Tyr Ile Ala Ile Asp Trp Glu Pro Glu Met Lys Lys	
65 70 75 80	
CGT TAC TAT GAC GAG GTA GAG GCT GAG GGC TAC GTG AAG CAT GAC TGC	288
Arg Tyr Tyr Asp Glu Val Glu Ala Glu Gly Tyr Val Lys His Asp Cys	
85 90 95	
GTC GGG TAC GTG ATG AAG AAG GCT CCC GTG CGG CTG CAG GAG TGC ATT	336
Val Gly Tyr Val Met Lys Lys Ala Pro Val Arg Leu Gln Glu Cys Ile	
100 105 110	
GAG CTC TTC ACC ACT GTG GAG ACC CTG GAG AAG GAA AAC CCC TGG TAC	384
Glu Leu Phe Thr Thr Val Glu Thr Leu Glu Lys Glu Asn Pro Trp Tyr	
115 120 125	
TGC CCT TCC TGC AAG CAG CAC CAG CTG GCA ACC AAG AAG CTG GAC CTG	432
Cys Pro Ser Cys Lys Gln His Gln Leu Ala Thr Lys Lys Leu Asp Leu	
130 135 140	
TGG ATG CTG CCG GAG ATT CTC ATC ATC CAC CTG AAA CGC TTT TCC TAC	480
Trp Met Leu Pro Glu Ile Leu Ile Ile His Leu Lys Arg Phe Ser Tyr	
145 150 155 160	
ACC AAG TTC TCC CGA GAG AAG CTG GAC ACC CTC GTG GAG TTT CCT ATC	528
Thr Lys Phe Ser Arg Glu Lys Leu Asp Thr Leu Val Glu Phe Pro Ile	
165 170 175	
CGG GAC CTG GAC TTC TCT GAG TTT GTC ATC CAG CCA CAG AAT GAG TCG	576
Arg Asp Leu Asp Phe Ser Glu Phe Val Ile Gln Pro Gln Asn Glu Ser	
180 185 190	
AAT CCG GAG CTG TAC AAA TAT GAC CTC ATC GCG GTT TCC AAC CAT TAT	624
Asn Pro Glu Leu Tyr Lys Tyr Asp Leu Ile Ala Val Ser Asn His Tyr	
195 200 205	
GGG GGC ATG CGT GAT GGA CAC TAC ACA ACA TTT GCC TGC AAC AAG GAC	672
Gly Gly Met Arg Asp Gly His Tyr Thr Thr Phe Ala Cys Asn Lys Asp	
210 215 220	
AGC GGC CAG TGG CAC TAC TTT GAT GAC AAC AGC GTC TCC CCT GTC AAT	720
Ser Gly Gln Trp His Tyr Phe Asp Asp Asn Ser Val Ser Pro Val Asn	
225 230 235 240	

GAG AAT CAG ATC GAG TCC AAG GCA GCC TAT GTC CTC TTC TAC CAA CGC	768
Glu Asn Gln Ile Glu Ser Lys Ala Ala Tyr Val Leu Phe Tyr Gln Arg	
245 250 255	
CAG GAC GTG GCG CGA CGC CTG CTG TCC CCG GCC GCC TCA TCT GGC GCC	816
Gln Asp Val Ala Arg Arg Leu Leu Ser Pro Ala Gly Ser Ser Gly Ala	
260 265 270	
CCAGCCTCCC CTGCCTGCAG CTCCCCACCC AGCTCTGAGT TCATGGATGT TAATTGAGAG	876
CCCTGGGGTCC TGCCACAGAA AA	898

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 272 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Leu Val Pro Arg Asp Pro Glu Pro Glu Gln Ala Gly Pro Ser Ser Gly  
1 5 10 15

Val Thr Asn Arg Cys Pro Phe Leu Leu Asp Asn Cys Leu Gly Thr Ser  
20 25 30

Gln Trp Pro Pro Arg Arg Arg Arg Lys Gln Leu Phe Thr Leu Gln Thr  
35 40 45

Val Asn Ser Asn Gly Thr Ser Asp Arg Thr Thr Ser Pro Glu Glu Val  
50 55 60

His Ala Gln Pro Tyr Ile Ala Ile Asp Trp Glu Pro Glu Met Lys Lys  
65 70 75 80

Arg Tyr Tyr Asp Glu Val Glu Ala Glu Gly Tyr Val Lys His Asp Cys  
85 90 95

Val Gly Tyr Val Met Lys Lys Ala Pro Val Arg Leu Gln Glu Cys Ile  
100 105 110

Glu Leu Phe Thr Thr Val Glu Thr Leu Glu Lys Glu Asn Pro Trp Tyr  
115 120 125

Cys Pro Ser Cys Lys Gln His Gln Leu Ala Thr Lys Lys Leu Asp Leu  
130 135 140

Trp Met Leu Pro Glu Ile Leu Ile Ile His Leu Lys Arg Phe Ser Tyr  
145 150 155 160

Thr Lys Phe Ser Arg Glu Lys Leu Asp Thr Leu Val Glu Phe Pro Ile  
165 170 175

Arg Asp Leu Asp Phe Ser Glu Phe Val Ile Gln Pro Gln Asn Glu Ser  
180 185 190

Asn Pro Glu Leu Tyr Lys Tyr Asp Leu Ile Ala Val Ser Asn His Tyr  
195 200 205

Gly Gly Met Arg Asp Gly His Tyr Thr Thr Phe Ala Cys Asn Lys Asp  
210 215 220

Ser Gly Gln Trp His Tyr Phe Asp Asp Asn Ser Val Ser Pro Val Asn  
 225 230 235 240  
 Glu Asn Gln Ile Glu Ser Lys Ala Ala Tyr Val Leu Phe Tyr Gln Arg  
 245 250 255  
 Gln Asp Val Ala Arg Arg Leu Leu Ser Pro Ala Gly Ser Ser Gly Ala  
 260 265 270

## (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

Pro Ser Trp Pro Glu Ser Lys Val Thr Glu Phe Leu His Gln Ser Lys  
 1 5 10 15  
 Leu Lys Ser Phe Glu Ser Glu Arg Val Gln Leu Leu Gln Glu Glu Thr  
 20 25 30  
 Ala Arg Asn Leu Thr Gln Cys Gln Leu Glu Cys Glu Lys Tyr Gln Lys  
 35 40 45  
 Lys Leu Glu Val Leu Thr Lys Glu Phe Tyr Ser Leu Gln Ala Ser Ser  
 50 55 60  
 Glu Lys Arg Ile Thr Glu Leu Gln Ala Gln Asn Ser Glu His Gln Ala  
 65 70 75 80  
 Arg Leu Asp Ile Tyr Glu Lys Leu Glu Lys Glu Leu Asp Glu Ile Ile  
 85 90 95  
 Met Gln Thr Ala Glu Ile Glu Asn Glu Asp Glu Ala Glu Arg Val Leu  
 100 105 110  
 Phe Ser Tyr Gly Tyr Gly Ala Asn Val Pro Thr Thr Ala Lys Arg Arg  
 115 120 125  
 Leu Lys Gln Ser Val His Leu Ala Arg Arg Val Leu Gln Leu Glu Lys  
 130 135 140  
 Gln Asn Ser Leu Ile Leu Lys Asp Leu Glu His Arg Lys Asp Gln Val  
 145 150 155 160  
 Thr Gln Leu Ser Gln Glu Leu Asp Arg Ala Asn Ser Leu Leu Asn Gln  
 165 170 175  
 Thr Gln Gln Pro Tyr Arg Tyr Leu Ile Glu Ser Val Arg Gln Arg Asp  
 180 185 190  
 Ser Lys Ile Asp Ser Leu Thr Glu Ser Ile Ala Gln Leu Glu Lys Asp  
 195 200 205  
 Val Ser Asn Leu Asn Lys Glu Lys Ser Ala Leu Leu Gln Thr Lys Asn  
 210 215 220  
 Gln Met Ala Leu Asp Leu Glu Gln Leu Leu Asn His Arg Glu Glu Leu  
 225 230 235 240

Ala	Ala	Met	Lys	Gln 245	Ile	Leu	Val	Lys	Met 250	His	Ser	Lys	His	Ser 255	Glu
Asn	Ser	Leu	Leu 260	Leu	Thr	Lys	Thr	Glu 265	Pro	Lys	His	Val	Thr 270	Glu	Asn
Gln	Lys	Ser 275	Lys	Thr	Leu	Asn	Val 280	Pro	Lys	Glu	His	Glu 285	Asp	Asn	Ile
Phe	Thr 290	Pro	Lys	Pro	Thr	Leu 295	Phe	Thr	Lys	Lys	Glu 300	Ala	Pro	Glu	Trp
Ser 305	Lys	Lys	Gln	Lys	Met 310	Lys	Thr								

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 587 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

Lys 1	Arg	Glu	Phe	Ile 5	Gln	Glu	Pro	Ala	Lys 10	Asn	Arg	Pro	Gly	Pro 15	Gln
Thr	Arg	Ser	Asp 20	Leu	Leu	Leu	Ser	Gly 25	Arg	Asp	Trp	Asn	Thr 30	Leu	Ile
Val	Gly	Lys 35	Leu	Ser	Pro	Trp	Ile 40	Arg	Pro	Asp	Ser	Lys 45	Val	Glu	Lys
Ile	Arg 50	Arg	Asn	Ser	Glu	Ala 55	Ala	Met	Leu	Gln	Glu 60	Leu	Asn	Phe	Gly
Ala 65	Tyr	Leu	Gly	Leu	Pro 70	Ala	Phe	Leu	Leu	Pro 75	Leu	Asn	Gln	Glu	Asp 80
Asn	Thr	Asn	Leu	Ala 85	Arg	Val	Leu	Thr	Asn 90	His	Ile	His	Thr	Gly 95	His
His	Ser	Ser	Met 100	Phe	Trp	Met	Arg	Val 105	Pro	Leu	Val	Ala	Pro 110	Glu	Asp
Leu	Arg	Asp 115	Asp	Ile	Ile	Glu	Asn 120	Ala	Pro	Thr	Thr	His 125	Thr	Glu	Glu
Tyr	Ser 130	Gly	Glu	Glu	Lys	Thr 135	Trp	Met	Trp	Trp	His 140	Asn	Phe	Arg	Thr
Leu 145	Cys	Asp	Tyr	Ser	Lys 150	Arg	Ile	Ala	Val	Ala 155	Leu	Glu	Ile	Gly	Ala 160
Asp	Leu	Pro	Ser	Asn 165	His	Val	Ile	Asp	Arg 170	Trp	Leu	Gly	Glu	Pro 175	Ile
Lys	Ala	Ala	Ile 180	Leu	Pro	Thr	Ser	Ile 185	Phe	Leu	Thr	Asn	Lys 190	Lys	Gly
Phe	Pro	Val 195	Leu	Ser	Lys	Met	His 200	Gln	Arg	Leu	Ile	Phe 205	Arg	Leu	Leu



Lys Leu Glu Val Gln Phe Ile Ile Thr Gly Thr Asn His His Ser Glu  
 210 215 220  
 Lys Glu Phe Cys Ser Tyr Leu Gln Tyr Leu Glu Tyr Leu Ser Gln Asn  
 225 230 235 240  
 Arg Pro Pro Pro Asn Ala Tyr Glu Leu Phe Ala Lys Gly Tyr Glu Asp  
 245 250 255  
 Tyr Leu Gln Ser Pro Leu Gln Pro Leu Met Asp Asn Leu Glu Ser Gln  
 260 265 270  
 Thr Tyr Glu Val Phe Glu Lys Asp Pro Ile Lys Tyr Ser Gln Tyr Gln  
 275 280 285  
 Gln Ala Ile Tyr Lys Cys Leu Leu Asp Arg Val Pro Glu Glu Glu Lys  
 290 295 300  
 Asp Thr Asn Val Gln Val Leu Met Val Leu Gly Ala Gly Arg Gly Pro  
 305 310 315 320  
 Leu Val Asn Ala Ser Leu Arg Ala Ala Lys Gln Ala Asp Arg Arg Ile  
 325 330 335  
 Lys Leu Tyr Ala Val Glu Lys Asn Pro Asn Ala Val Val Thr Leu Glu  
 340 345 350  
 Asn Trp Gln Phe Glu Glu Trp Gly Ser Gln Val Thr Val Val Ser Ser  
 355 360 365  
 Asp Met Arg Glu Trp Val Ala Pro Glu Lys Ala Asp Ile Ile Val Ser  
 370 375 380  
 Glu Leu Leu Gly Ser Phe Ala Asp Asn Glu Leu Ser Pro Glu Cys Leu  
 385 390 395 400  
 Asp Gly Ala Gln His Phe Leu Lys Asp Asp Gly Val Ser Ile Pro Gly  
 405 410 415  
 Glu Tyr Thr Ser Phe Leu Ala Pro Ile Ser Ser Ser Lys Leu Tyr Asn  
 420 425 430  
 Glu Val Arg Ala Cys Arg Glu Lys Asp Arg Asp Pro Glu Ala Gln Phe  
 435 440 445  
 Glu Met Pro Tyr Val Val Arg Leu His Asn Phe His Gln Leu Ser Ala  
 450 455 460  
 Pro Gln Pro Cys Phe Thr Phe Ser His Pro Asn Arg Asp Pro Met Ile  
 465 470 475 480  
 Asp Asn Asn Arg Tyr Cys Thr Leu Glu Phe Pro Val Glu Val Asn Thr  
 485 490 495  
 Val Leu His Gly Phe Ala Gly Tyr Phe Glu Thr Val Leu Tyr Gln Asp  
 500 505 510  
 Ile Thr Leu Ser Ile Arg Pro Glu Thr His Ser Pro Gly Met Phe Ser  
 515 520 525  
 Trp Phe Pro Ile Leu Phe Pro Ile Lys Gln Pro Ile Thr Val Arg Glu  
 530 535 540  
 Gly Gln Thr Ile Cys Val Arg Phe Trp Arg Cys Ser Asn Ser Lys Lys  
 545 550 555 560

Val Trp Tyr Glu Trp Ala Val Thr Ala Pro Val Cys Ser Ala Ile His  
565 570 575

Asn Pro Thr Gly Arg Ser Tyr Thr Ile Gly Leu  
580 585

## (2) INFORMATION FOR SEQ ID NO: 50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 370 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 110..111
- (D) OTHER INFORMATION: /note= "Xaa = Glu or Lys"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Glu Pro Gly Arg Gly Leu Leu Val Ser Val Met Ala His Glu Ala Met  
1 5 10 15  
Glu Tyr Asp Val Gln Val Gln Leu Asn His Ala Glu Gln Gln Pro Ala  
20 25 30  
Pro Ala Gly Met Ala Ser Ser Gln Gly Gly Pro Ala Leu Leu Gln Pro  
35 40 45  
Val Pro Ala Asp Val Val Ser Ser Gln Gly Val Pro Ser Ile Leu Gln  
50 55 60  
Pro Ala Pro Ala Glu Val Ile Ser Ser Gln Ala Thr Pro Pro Leu Leu  
65 70 75 80  
Gln Pro Ala Pro Gln Leu Ser Val Asp Leu Thr Glu Val Glu Val Leu  
85 90 95  
Gly Glu Asp Thr Val Glu Asn Ile Asn Pro Arg Thr Ser Xaa Gln His  
100 105 110  
Arg Gln Gly Ser Asp Gly Asn His Thr Ile Pro Ala Ser Ser Leu His  
115 120 125  
Ser Met Thr Asn Phe Ile Ser Gly Leu Gln Arg Leu His Gly Met Leu  
130 135 140  
Glu Phe Leu Arg Pro Ser Ser Ser Asn His Ser Val Gly Pro Met Arg  
145 150 155 160  
Thr Arg Arg Arg Val Ser Ala Ser Arg Arg Ala Arg Ala Gly Gly Ser  
165 170 175  
Gln Arg Thr Asp Ser Ala Arg Leu Arg Ala Pro Leu Asp Ala Tyr Phe  
180 185 190  
Gln Val Ser Arg Thr Gln Pro Asp Leu Pro Ala Thr Thr Tyr Asp Ser  
195 200 205  
Glu Thr Arg Asn Pro Val Ser Glu Glu Leu Gln Val Ser Ser Ser Ser  
210 215 220

285

Asp Ser Asp Ser Asp Ser Ser Ala Glu Tyr Gly Gly Val Val Asp Gln  
 225 230 235 240  
 Ala Glu Glu Ser Gly Ala Val Ile Leu Glu Glu Gln Leu Ala Gly Val  
 245 250 255  
 Ser Ala Glu Gln Glu Val Thr Cys Ile Asp Gly Gly Lys Thr Leu Pro  
 260 265 270  
 Lys Gln Pro Ser Pro Gln Lys Ser Glu Pro Leu Leu Pro Ser Ala Ser  
 275 280 285  
 Met Asp Glu Glu Glu Gly Asp Thr Cys Thr Ile Cys Leu Glu Gln Trp  
 290 295 300  
 Thr Asn Ala Gly Asp His Arg Leu Ser Ala Leu Arg Cys Gly His Leu  
 305 310 315 320  
 Phe Gly Tyr Arg Cys Ile Ser Thr Trp Leu Lys Gly Gln Val Arg Lys  
 325 330 335  
 Cys Pro Gln Cys Asn Lys Lys Ala Arg His Ser Asp Ile Val Val Leu  
 340 345 350  
 Tyr Ala Arg Thr Leu Arg Ala Leu Asp Thr Ser Glu Gln Glu Arg Met  
 355 360 365  
 Lys Arg  
 370

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 416 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

Asp Tyr His Gln Asn Trp Gly Arg Asp Gly Gly Pro Arg Ser Ser Gly  
 1 5 10 15  
 Gly Gly Tyr Gly Gly Gly Pro Ala Gly Gly His Gly Gly Asn Arg Gly  
 20 25 30  
 Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly Arg Gly Gly Arg Gly  
 35 40 45  
 Arg His Pro Gly His Leu Lys Gly Arg Glu Ile Gly Met Trp Tyr Ala  
 50 55 60  
 Lys Lys Gln Gly Gln Lys Asn Lys Glu Ala Glu Arg Gln Glu Arg Ala  
 65 70 75 80  
 Val Val His Met Asp Glu Arg Arg Glu Glu Gln Ile Val Gln Leu Leu  
 85 90 95  
 Asn Ser Val Gln Ala Lys Asn Asp Lys Glu Ser Glu Ala Gln Ile Ser  
 100 105 110  
 Trp Phe Ala Pro Glu Asp His Gly Tyr Gly Thr Glu Val Ser Thr Lys  
 115 120 125

Asn Thr Pro Cys Ser Glu Asn Lys Leu Asp Ile Gln Glu Lys Lys Leu  
 130 135 140  
 Ile Asn Gln Glu Lys Lys Met Phe Arg Ile Arg Asn Arg Ser Tyr Ile  
 145 150 155 160  
 Asp Arg Asp Ser Glu Tyr Leu Leu Gln Glu Asn Glu Pro Asp Gly Thr  
 165 170 175  
 Leu Asp Gln Lys Leu Leu Glu Asp Leu Gln Lys Lys Lys Asn Asp Leu  
 180 185 190  
 Arg Tyr Ile Glu Met Gln His Phe Arg Glu Lys Leu Pro Ser Tyr Gly  
 195 200 205  
 Met Gln Lys Glu Leu Val Asn Leu Ile Asp Asn His Gln Val Thr Val  
 210 215 220  
 Ile Ser Gly Glu Thr Gly Cys Gly Lys Thr Thr Gln Val Thr Gln Phe  
 225 230 235 240  
 Ile Leu Asp Asn Tyr Ile Glu Arg Gly Lys Gly Ser Ala Cys Arg Ile  
 245 250 255  
 Val Cys Thr Gln Pro Arg Arg Ile Ser Ala Ile Ser Val Ala Glu Arg  
 260 265 270  
 Val Ala Ala Glu Arg Ala Glu Ser Cys Gly Ser Gly Asn Ser Thr Gly  
 275 280 285  
 Tyr Gln Ile Arg Leu Gln Ser Arg Leu Pro Arg Lys Gln Gly Ser Ile  
 290 295 300  
 Leu Tyr Cys Thr Thr Gly Ile Ile Leu Gln Trp Leu Gln Ser Asp Pro  
 305 310 315 320  
 Tyr Leu Ser Ser Val Ser His Ile Val Leu Asp Glu Ile His Glu Arg  
 325 330 335  
 Asn Leu Gln Ser Asp Val Leu Met Thr Val Val Lys Asp Leu Leu Asn  
 340 345 350  
 Phe Arg Ser Asp Leu Lys Val Ile Leu Met Ser Ala Thr Leu Asn Ala  
 355 360 365  
 Glu Lys Phe Ser Glu Tyr Phe Gly Asn Cys Pro Met Ile His Ile Pro  
 370 375 380  
 Gly Phe Thr Phe Pro Val Val Glu Tyr Leu Leu Glu Asp Val Ile Glu  
 385 390 395 400  
 Lys Ile Arg Tyr Val Pro Glu Gln Lys Glu His Arg Ser Gln Phe Lys  
 405 410 415

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```

Asn Ile Ser Trp Lys Lys Thr Ile Val Thr Arg Phe Leu Lys Leu Val
1           5           10           15
Pro Asp Leu Leu Ala Ile Val Gln Arg Lys Lys Lys Glu Gly Glu Glu
20           25           30
Glu Gln Ala Ile Asn Arg Gln Thr Ala Leu Tyr Thr Leu Lys Leu Leu
35           40           45
Cys Lys Asn Phe Gly Ala Glu Asn Pro Asp Pro Phe Val Pro Val Leu
50           55           60
Ser Thr Ala Val Lys Leu Ile Ala Pro Glu Arg Lys Glu Glu Lys Asn
65           70           75           80
Val Leu Gly Ser Ala Leu Leu Cys Met Ala Glu Val Thr Ser Thr Leu
85           90           95
Glu Ala Leu Ala Ile Pro Gln Leu Pro Ser Leu Met Pro Ser Leu Leu
100          105          110
Thr Thr Met Lys Asn Thr Ser Glu Leu Val Ser Ser Glu Val Tyr Leu
115          120          125
Leu Ser Ala Leu Ala Ala Leu Gln Lys Val Val Glu Thr Leu Pro His
130          135          140
Phe Ile Ser Pro Tyr Leu Glu Gly Ile Leu Ser Gln Val Ile His Leu
145          150          155          160
Glu Lys Ile Thr Ser Glu Met Gly Ser Ala Ser Gln Ala Asn Ile Arg
165          170          175
Leu Thr Ser Leu Lys Lys Thr Leu Ala Thr Thr Leu Ala Pro Arg Val
180          185          190
Leu Leu Pro Ala Ile Lys Lys Thr Tyr Lys Gln Ile Glu Lys Asn Trp
195          200          205
Lys Asn His Met Gly Pro Phe Met Ser Ile Leu Gln Glu His Ile Gly
210          215          220
Ala Met Lys Lys Glu Glu Leu Thr Ser His Gln Ser Gln Leu Thr Ala
225          230          235          240
Phe Phe Leu Glu Ala Leu Asp Phe Arg Ala Gln His Ser Glu Asn Asp
245          250          255
Leu Glu Glu Val Gly Lys Thr Glu Asn Cys Ile Ile Asp Cys Leu Val
260          265          270
Ala Met Val Val Lys Leu Ser Glu Val Thr Phe Arg Pro Leu Phe Phe
275          280          285
Lys Leu Phe Asp Trp Ala Lys Thr Glu Asp Ala Pro Lys Asp Arg Leu
290          295          300
Leu Thr Phe Tyr Asn Leu Ala Asp Cys Ile Ala Glu Lys Leu Lys Gly
305          310          315          320
Leu Phe Thr Leu Phe Ala Gly His Leu Val Lys Pro Phe Ala Asp Thr
325          330          335

```

Leu Asp Gln Val Asn Ile Ser Lys Thr Asp Glu Ala Phe Phe Asp Ser  
 340 345 350  
 Glu Asn Asp Pro Glu Lys Cys Cys Leu Leu Leu Gln Phe Ile Leu Asn  
 355 360 365  
 Cys Leu Tyr Lys Ile Phe Leu Phe Asp Thr Gln His Phe Ile Ser Lys  
 370 375 380  
 Glu Arg Ala Gly Ala Leu Met Met Pro Leu Val Asp Gln Leu Glu Asn  
 385 390 395 400  
 Arg Leu Gly Gly Glu Glu Lys Phe Gln Glu Arg Val Thr Lys His Leu  
 405 410 415  
 Ile Pro Cys Ile Ala Gln Phe Ser Val Ala Met Ala Asp Asp Ser Leu  
 420 425 430  
 Trp Lys Pro Leu Asn Tyr Gln Ile Leu Leu Lys Thr Arg Asp Ser Ser  
 435 440 445  
 Pro Lys Val Arg Phe Ala Ala Leu Ile Thr Val Leu Ala Leu Ala Glu  
 450 455 460  
 Lys Leu Lys Glu Asn Tyr Ile Val Leu Leu Pro Glu Ser Ile Pro Phe  
 465 470 475 480  
 Leu Ala Glu Leu Met Glu Asp Glu Cys Glu Glu Val Glu His Gln Cys  
 485 490 495  
 Gln Lys Thr Ile Gln Gln Leu Glu Thr Val Leu Gly Glu Pro Leu Gln  
 500 505 510  
 Ser Tyr Phe  
 515

## (2) INFORMATION FOR SEQ ID NO: 53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 149 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

Gly Val Val Pro Asn Gly Arg Asp Ala Glu Ser Gly His Ser Leu Ala  
 1 5 10 15  
 Glu Gly Gln Ala Pro His Gly Leu Pro Gly Thr Pro Gly Ala Ser Gly  
 20 25 30  
 Gly Val Val Leu Gln Pro Arg Gly Arg Arg Arg Ala Asp Pro Pro His  
 35 40 45  
 Arg Gln Leu Arg Pro Glu Ala Phe Gly Asn His Arg Arg Ser Glu Phe  
 50 55 60  
 Leu Arg Leu Gln Val Glu Gly Gly Gly Cys Ser Gly Phe Gln Tyr Lys  
 65 70 75 80  
 Phe Ser Leu Asp Thr Val Ile Asn Pro Asp Asp Arg Val Phe Glu Gln  
 85 90 95

Gly Gly Ala Arg Val Val Val Asp Ser Asp Ser Leu Ala Phe Val Lys  
                           100                          105                          110

Gly Ala Gln Val Asp Phe Ser Gln Glu Leu Ile Arg Ser Ser Phe Gln  
                   115                          120                          125

Val Leu Asn Asn Pro Gln Ala Gln Gln Gly Cys Ser Cys Gly Ser Ser  
           130                          135                          140

Phe Ser Ile Lys Leu  
           145

## (2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 535 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

Gly Pro Ala Gly Gly Ala Pro Thr Pro Ala Leu Val Ala Gly Ser Ser  
   1                          5                          10                          15

Ala Ala Ala Pro Phe Pro His Gly Asp Ser Ala Leu Asn Glu Gln Glu  
                   20                          25                          30

Lys Glu Leu Gln Arg Arg Leu Lys Arg Leu Tyr Pro Ala Val Asp Glu  
           35                          40                          45

Gln Glu Thr Pro Leu Pro Arg Ser Trp Ser Pro Lys Asp Lys Phe Ser  
           50                          55                          60

Tyr Ile Gly Leu Ser Gln Asn Asn Leu Arg Val His Tyr Lys Gly His  
   65                          70                          75                          80

Gly Lys Thr Pro Lys Asp Ala Ala Ser Val Arg Ala Thr His Pro Ile  
                   85                          90                          95

Pro Ala Ala Cys Gly Ile Tyr Tyr Phe Glu Val Lys Ile Val Ser Lys  
                   100                          105                          110

Gly Arg Asp Gly Tyr Met Gly Ile Gly Leu Ser Ala Gln Gly Val Asn  
           115                          120                          125

Met Asn Arg Leu Pro Gly Trp Asp Lys His Ser Tyr Gly Tyr His Gly  
   130                          135                          140

Asp Asp Gly His Ser Phe Cys Ser Ser Gly Thr Gly Gln Pro Tyr Gly  
   145                          150                          155                          160

Pro Thr Phe Thr Thr Gly Asp Val Ile Gly Cys Cys Val Asn Leu Ile  
                   165                          170                          175

Asn Asn Thr Cys Phe Tyr Thr Lys Asn Gly His Ser Leu Gly Ile Ala  
                   180                          185                          190

Phe Thr Asp Leu Pro Pro Asn Leu Tyr Pro Thr Val Gly Leu Gln Thr  
           195                          200                          205

Pro Gly Glu Val Val Asp Ala Asn Phe Gly Gln His Pro Phe Val Phe  
   210                          215                          220

Asp Ile Glu Asp Tyr Met Arg Glu Trp Arg Thr Lys Ile Gln Ala Gln  
 225 230 235 240  
 Ile Asp Arg Phe Pro Ile Gly Asp Arg Glu Gly Glu Trp Gln Thr Met  
 245 250 255  
 Ile Gln Lys Met Val Ser Ser Tyr Leu Val His His Gly Tyr Cys Ala  
 260 265 270  
 Thr Ala Glu Ala Phe Ala Arg Ser Thr Asp Gln Thr Val Leu Glu Glu  
 275 280 285  
 Leu Ala Ser Ile Lys Asn Arg Gln Arg Ile Gln Lys Leu Val Leu Ala  
 290 295 300  
 Gly Arg Met Gly Glu Ala Ile Glu Thr Thr Gln Gln Leu Tyr Pro Ser  
 305 310 315 320  
 Leu Leu Glu Arg Asn Pro Asn Leu Leu Phe Thr Leu Lys Val Arg Gln  
 325 330 335  
 Phe Ile Glu Met Val Asn Gly Thr Asp Ser Glu Val Arg Cys Leu Gly  
 340 345 350  
 Gly Arg Ser Pro Lys Ser Gln Asp Ser Tyr Pro Val Ser Pro Arg Pro  
 355 360 365  
 Phe Ser Ser Pro Ser Met Ser Pro Ser His Gly Met Asn Ile His Asn  
 370 375 380  
 Leu Ala Ser Gly Lys Gly Ser Thr Ala His Phe Ser Gly Phe Glu Ser  
 385 390 395 400  
 Cys Ser Asn Gly Val Ile Ser Asn Lys Ala His Gln Ser Tyr Cys His  
 405 410 415  
 Ser Asn Lys His Gln Ser Ser Asn Leu Asn Val Pro Glu Leu Asn Ser  
 420 425 430  
 Ile Asn Met Ser Arg Ser Gln Gln Val Asn Asn Phe Thr Ser Asn Asp  
 435 440 445  
 Val Asp Met Glu Thr Asp His Tyr Ser Asn Gly Val Gly Glu Thr Ser  
 450 455 460  
 Ser Asn Gly Phe Leu Asn Gly Ser Ser Lys His Asp His Glu Met Glu  
 465 470 475 480  
 Asp Cys Asp Thr Glu Met Glu Val Asp Ser Ser Gln Leu Arg Arg Gln  
 485 490 495  
 Leu Cys Gly Gly Ser Gln Ala Ala Ile Glu Arg Met Ile His Phe Gly  
 500 505 510  
 Arg Glu Leu Gln Ala Met Ser Glu Gln Leu Arg Arg Asp Cys Gly Lys  
 515 520 525  
 Asn Thr Ala Asn Lys Lys Cys  
 530 535



## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 395 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

Val Val Lys Pro Pro Gly Ser Ser Leu Asn Gly Val His Pro Asn Pro  
 1 5 10 15  
 Thr Pro Ile Val Gln Arg Leu Pro Ala Phe Leu Asp Asn His Asn Tyr  
 20 25 30  
 Ala Lys Ser Pro Met Gln Glu Glu Glu Asp Leu Ala Ala Gly Val Gly  
 35 40 45  
 Arg Ser Arg Val Pro Val Arg Pro Pro Gln Gln Tyr Ser Asp Asp Glu  
 50 55 60  
 Asp Asp Tyr Glu Asp Asp Glu Glu Asp Asp Val Gln Asn Thr Asn Ser  
 65 70 75 80  
 Ala Leu Arg Tyr Lys Gly Lys Gly Thr Gly Lys Pro Gly Ala Leu Ser  
 85 90 95  
 Gly Ser Ala Asp Gly Gln Leu Ser Val Leu Gln Pro Asn Thr Ile Asn  
 100 105 110  
 Val Leu Ala Glu Lys Leu Lys Glu Ser Gln Lys Asp Leu Ser Ile Pro  
 115 120 125  
 Leu Ser Ile Lys Thr Ser Ser Gly Ala Gly Ser Pro Ala Val Ala Val  
 130 135 140  
 Pro Thr His Ser Gln Pro Ser Pro Thr Pro Ser Asn Glu Ser Thr Asp  
 145 150 155 160  
 Thr Ala Ser Glu Ile Gly Ser Ala Phe Asn Ser Pro Leu Arg Ser Pro  
 165 170 175  
 Ile Arg Ser Ala Asn Pro Thr Arg Pro Ser Ser Pro Val Thr Ser His  
 180 185 190  
 Ile Ser Lys Val Leu Phe Gly Glu Asp Asp Ser Leu Leu Arg Val Asp  
 195 200 205  
 Cys Ile Arg Tyr Asn Arg Ala Val Arg Asp Leu Gly Pro Val Ile Ser  
 210 215 220  
 Thr Gly Leu Leu His Leu Ala Glu Asp Gly Val Leu Ser Pro Leu Ala  
 225 230 235 240  
 Leu Thr Glu Gly Gly Lys Gly Ser Ser Pro Ser Ile Arg Pro Ile Gln  
 245 250 255  
 Gly Ser Gln Gly Ser Ser Ser Pro Val Glu Lys Glu Val Val Glu Ala  
 260 265 270  
 Thr Asp Ser Arg Glu Lys Thr Gly Met Val Arg Pro Gly Glu Pro Leu  
 275 280 285

Ser Gly Glu Lys Tyr Ser Pro Lys Glu Leu Leu Ala Leu Leu Lys Cys  
 290 295 300  
 Val Glu Ala Glu Ile Ala Asn Tyr Glu Ala Cys Leu Lys Glu Glu Val  
 305 310 315 320  
 Glu Lys Arg Lys Lys Phe Lys Ile Asp Asp Gln Arg Arg Thr His Asn  
 325 330 335  
 Tyr Asp Glu Phe Ile Cys Thr Phe Ile Ser Met Leu Ala Gln Glu Gly  
 340 345 350  
 Met Leu Ala Asn Leu Val Glu Gln Asn Ile Ser Val Arg Arg Arg Gln  
 355 360 365  
 Gly Ala Ser Ile Gly Arg Leu His Lys Gln Arg Lys Pro Asp Arg Arg  
 370 375 380  
 Lys Arg Ser Arg Pro Tyr Lys Ala Lys Arg Gln  
 385 390 395

## (2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 278 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Met Val Lys Val Lys Gly Gln Val Ser Glu Met Ala Val Leu Leu Ile  
 1 5 10 15  
 Asp Pro Glu Pro Gln Ile Ala Ala Leu Ala Lys Asn Phe Phe Asn Glu  
 20 25 30  
 Leu Ser His Lys Gly Asn Ala Ile Tyr Asn Leu Leu Pro Asp Ile Ile  
 35 40 45  
 Ser Arg Leu Ser Asp Pro Glu Leu Gly Val Glu Glu Glu Pro Phe His  
 50 55 60  
 Thr Ile Met Lys Gln Leu Leu Ser Tyr Ile Thr Lys Asp Lys Gln Thr  
 65 70 75 80  
 Glu Ser Leu Val Glu Lys Leu Cys Gln Arg Phe Arg Thr Ser Leu Thr  
 85 90 95  
 Glu Arg Gln Gln Arg Asp Leu Ala Tyr Cys Val Ser Gln Leu Pro Leu  
 100 105 110  
 Thr Glu Arg Gly Leu Arg Lys Met Leu Asp Asn Phe Asp Cys Phe Gly  
 115 120 125  
 Asp Lys Leu Ser Asp Glu Ser Ile Phe Ser Ala Phe Leu Ser Val Val  
 130 135 140  
 Gly Lys Leu Arg Arg Gly Ala Lys Pro Glu Gly Lys Ala Ile Ile Asp  
 145 150 155 160  
 Glu Phe Glu Gln Lys Leu Arg Ala Cys His Thr Arg Gly Leu Asp Gly  
 165 170 175

293

Ile Lys Glu Leu Glu Ile Gly Gln Ala Gly Ser Gln Arg Ala Pro Ser  
 180 185 190

Ala Lys Lys Pro Ser Thr Gly Ser Arg Tyr Gln Pro Leu Ala Ser Thr  
 195 200 205

Ala Ser Asp Asn Asp Phe Val Thr Pro Glu Pro Arg Arg Thr Thr Arg  
 210 215 220

Arg His Pro Asn Thr Gln Gln Arg Ala Ser Lys Lys Lys Pro Lys Val  
 225 230 235 240

Val Phe Ser Ser Asp Glu Ser Ser Glu Glu Asp Leu Ser Ala Glu Met  
 245 250 255

Thr Glu Asp Glu Thr Pro Lys Lys Thr Thr Pro Ile Leu Arg Ala Ser  
 260 265 270

Ala Arg Arg His Arg Ser  
 275

## (2) INFORMATION FOR SEQ ID NO: 57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GCGAGGAGCC TTTCATCCGA

20

## (2) INFORMATION FOR SEQ ID NO: 58:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGAGCGCGGC GCGACTGT

18

## (2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

ATGGAACCGG ATGGTCGCGG T

21

## (2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TCTTCAAGTC TTGTATCCAG GC

22

## (2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CGCCATGGAA CCAAATACA

19

## (2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

GCCTGGATAC AAGACTTGAA G

21

## (2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTGTAGACGT CCTCCTGAAC C

21

## (2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

AAAGCTTCAG TGCAAACCCA

20

## (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TCCAGATCTT GCAGAAGCC

19

## (2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CAGATGTTTC TGAGAGGGCT

20

## (2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

ATTCCTCTTT GGAGTCAAAT TC

22

## (2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

GAGGCAGAAA AAGAAGATGG T

21

## (2) INFORMATION FOR SEQ ID NO: 69:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AGGAGCCACT TGCTAGTAAG

20

## (2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

ATGGTGAAAT AGACTTACTA GC

22

## (2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

GCAGACCTTC TCAGGAGTC

19

## (2) INFORMATION FOR SEQ ID NO: 72:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

AAGAGCAGGA ATGAAGTAGT G

21

## (2) INFORMATION FOR SEQ ID NO: 73:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CTCCACTGGT GCTCAGAATG

20

## (2) INFORMATION FOR SEQ ID NO: 74:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

AGTGGAGATT TTGTTAAGCA A

21

## (2) INFORMATION FOR SEQ ID NO: 75:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

AGGTGGTGTA GGTGGTGAA

19

## (2) INFORMATION FOR SEQ ID NO: 76:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

GGTACACCAC CTTCTACATT

20

## (2) INFORMATION FOR SEQ ID NO: 77:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

GTCTCTCCTC TATGATTTCT T

21

## (2) INFORMATION FOR SEQ ID NO: 78:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CAATGAAGCT GTTGCCCAA

19

## (2) INFORMATION FOR SEQ ID NO: 79:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GTCTTTAACA TTTGGATCAC T

21

## (2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 21 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

AGTGATCCAA ATGTAAAGA C

21

## (2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 21 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

AGTGATCCAA ATGTAAAGA C

21

## (2) INFORMATION FOR SEQ ID NO: 82:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 21 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

CAAAATGACT CACCACTTCA C

21

## (2) INFORMATION FOR SEQ ID NO: 83:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 18 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

ATCGACAGGC CGCAGACC

18

## (2) INFORMATION FOR SEQ ID NO: 84:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 21 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGTCGATT ATACAGATGA T

21



## (2) INFORMATION FOR SEQ ID NO: 85:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

AACATGAGTT ACTGTACTGT C

21

## (2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TATACTGAGT TTGACAGTAC AG

22

## (2) INFORMATION FOR SEQ ID NO: 87:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CATACTTTTC TTCGTAGACA TG

22

## (2) INFORMATION FOR SEQ ID NO: 88:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

TGGGTAAAAG CATGTCTACG A

21

## (2) INFORMATION FOR SEQ ID NO: 89:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GGATGCTACT TCTATTTGTG

20

## (2) INFORMATION FOR SEQ ID NO: 90:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GAGTCACGTC ACTGTCTG

18

## (2) INFORMATION FOR SEQ ID NO: 91:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

CCTCAGTAGA AAGCCCAAGC

20

## (2) INFORMATION FOR SEQ ID NO: 92:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GCCCCTGCCG AACCTCTC

19

## (2) INFORMATION FOR SEQ ID NO: 93:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

GAGAGGGTTC GGCAGGGC

18

## (2) INFORMATION FOR SEQ ID NO: 94:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

TTCAATTTC AATGTTTCATC TGGT

24

## (2) INFORMATION FOR SEQ ID NO: 95:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

ACAGTCGCGC CGCGCTCGA

19

## (2) INFORMATION FOR SEQ ID NO: 96:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

CAGAAACTGT GCGACCCGTG

20

## (2) INFORMATION FOR SEQ ID NO: 97:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

AGATGTTTAT CTAACAATGA CTC

23

## (2) INFORMATION FOR SEQ ID NO: 98:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

AGTTGTACTA TATACATCAA ACC

23

## (2) INFORMATION FOR SEQ ID NO: 99:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

ATTCTGCTGA ATGGGTTGCT T

21

## (2) INFORMATION FOR SEQ ID NO: 100:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

TAACTAAGAG AGATAGGGAT AG

22

## (2) INFORMATION FOR SEQ ID NO: 101:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

GGAGCTCCAT GTGGGAGCAA

20

## (2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

AACATCTGCA GGAGGACTTG G

21

## (2) INFORMATION FOR SEQ ID NO: 103:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

TCTGAGATGG TATTCAGAG T

21

## (2) INFORMATION FOR SEQ ID NO: 104:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

TGCTTTTAA TTTCCATTT GTTC

24

## (2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

AAGAACTGTA AAACACAGAA AGA

23

## (2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

TGCTCTTTCT TATCACTTCT TTC

23

## (2) INFORMATION FOR SEQ ID NO: 107:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

CTTGACTCAA GAATATAGGT CC

22

## (2) INFORMATION FOR SEQ ID NO: 108:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

TAGTGCTCAC TTGATACTTA GT

22

## (2) INFORMATION FOR SEQ ID NO: 109:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CATAATAAGA ACAATGAAAG TTGT

24

## (2) INFORMATION FOR SEQ ID NO: 110:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TTGATCTGCC TTTAACAAAT G

21

## (2) INFORMATION FOR SEQ ID NO: 111:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 153 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

Gly Leu Leu Ala Leu Thr Leu Gln Pro Thr Leu Ala Val Trp Pro Ser  
 1 5 10 15  
 Pro Gly Ser Phe Pro Ala Pro Leu Pro Leu Phe Pro Val Leu Leu Asn  
 20 25 30  
 Ser Pro Ser Trp Arg Val Gln Ala Leu Gly Met Gly Gly Thr Arg Pro  
 35 40 45  
 His Ser Phe His Arg Ala Leu Arg Pro Asp Thr Ala Asp Gln Pro His  
 50 55 60  
 Ser Ala Gln Glu Ala Ala Ser Gly Val Gly Ala Gln Arg Gly Thr Ala  
 65 70 75 80  
 Ala Ser Ser Thr Ala Gly Cys Gly Ala Ala Gly Pro Gly Pro Ser Ala  
 85 90 95  
 Trp Ala Ala Glu Tyr Ile Phe Tyr Leu Ser Glu Thr Ser Ile Phe Leu  
 100 105 110  
 Gly Ser Asn Pro Thr Cys His His Val Asp Ile Ser Ser Tyr Leu Thr  
 115 120 125  
 Met Leu Ser Leu Leu Arg Ser Cys Pro Gly Gly Pro Arg Ser Leu Tyr  
 130 135 140  
 His Ala Thr Val Pro Thr Thr Gly Ser  
 145 150

## (2) INFORMATION FOR SEQ ID NO: 112:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

TACCCTATAA GCCAGAATCC A

21

## (2) INFORMATION FOR SEQ ID NO: 113:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

GGCAAACCTTG TACACGAGCA

20

## (2) INFORMATION FOR SEQ ID NO: 114:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

GGTACTAGTG AAATCACCAG T

21

## (2) INFORMATION FOR SEQ ID NO: 115:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

GTGAATGCGT GCTACATTCA T

21

## (2) INFORMATION FOR SEQ ID NO: 116:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TTGAGTCGAG TCACACATTT GA

22

## (2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CTATTATGTT CCTTTCATAA CCA

23

## (2) INFORMATION FOR SEQ ID NO: 118:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

TAATGTCTTT GTCTAGTCGT CTAA

24

## (2) INFORMATION FOR SEQ ID NO: 119:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

GGTAGTTCTC CAAAAGGATC A

21

## (2) INFORMATION FOR SEQ ID NO: 120:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

GAGTTATAAG AAGCAGGCCA A

21

## (2) INFORMATION FOR SEQ ID NO: 121:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

ATTTCTTAAT TCTCTCAAAT CCAA

24

## (2) INFORMATION FOR SEQ ID NO: 122:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2856 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
(B) LOCATION: 233  
(D) OTHER INFORMATION: /note= "H = A, C or T"



## (ix) FEATURE:

- (A) NAME/KEY: modified\_base  
 (B) LOCATION:359  
 (D) OTHER INFORMATION:/note= "Y = C or T"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature  
 (B) LOCATION:2031..2188  
 (D) OTHER INFORMATION:/note= "Exon I"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

TGCCCTCATA ACCCATCTA AATTTAACTA CCACCCAAAG GTCCCCTCCA AATACTATCA	60
CGTTGGGGTT ACAACTTTTA ATATATGAAT TTGGGGGTGA CACTACAGAT ACTAATGCCC	120
ATTTTCATAGG GTCCCTATAA GGCTTAAGGC AGGTATTAAC ATAGGAAAGC ACTTAAAGCT	180
GGGTCTGGCT TGGGTAGGTA GTTCAATGAT TCAACAAACA CTGAGCACCT ACHTGGAGCC	240
AAGCACTGCA TGTGCCACAT GAAGCGATAT TGGGAAATGA GTCACATGCA GCCAATCTCT	300
GGCCTTTTGG AGGTTTTGAA CTAGAAGGGG ACACGCACAT AATCGTATGT GTGTGTATYT	360
ACATACACAG GTGATATGAT GCACCTGAGA GAATCCAGTC TAGGAACTAG GAAAACCTTT	420
AAGGAGTGAT ACTTCAGCTG TATTCTGAAG GATGAGGAAT GGAGAGGAGG CCAATTCCAG	480
GTTCCAAAGT GAATCTTTGC GCAAAAGCCA TGAGGCAGCA AGGTGCAGGG GCTTTTAATG	540
ACCTAAGGGA ACGCACTGTG GGGTTGGGAC CATGATGGCC AGAGGAGGTT TTGACAAGAG	600
GCTAGTCAAA GAGCAGAGAA AAACCTTTAAG GAGTTTTAAG AAAGGGAAGT GCCACGATGA	660
GTTTTGTGTT TGGAAATGTT TTAGGCGGCC ACACCGCAGT CTCGGGACTG GCTGGGACCT	720
GGATAGACAC TTGGATATCA GCTAGAAAGC TACGACAGGA AACCAGGCGA GGACGAGGCA	780
ACTGGGGATG TTGTGGAGCA GAGACGGAGA GAAAATGGGT GCATTCGAGA CAAGTTAGGG	840
GGAAAAAATG CAAGGACTTG GAAATGAACT TGGGGCGCGG CAGGAAGGCA TGACGGGTTG	900
CTCTGTAGGT CTTATCTGTA AATTACGGCG ATCAGTGAAA GATCTGGAGG AGGAAGGTGG	960
ACACACTCTC TAACAAAAAA AACCCTTTTT GAAATTTTAT ACCAATATTT TAAAAGTAAA	1020
CCAGATCTTT TCAGACATGC CTTTGAGCTG ATATTTGTTA ACTAGTTAGA ATTAGAAACT	1080
TTCCTTATTT TTAAGTCACT ACAATATACG CCACAGCTGA GGTGAGAGGA AAGAAAAGGT	1140
TGCTTTCTTA GGAACAAAGA GTGGTACCTT CAGTATCGTG GGCAAAGCTT TTCCAAGTCC	1200
AACAGTAGTC AAAACAGCGC TTTTATATAA TAACACTCAG CTAAAAGTTT CTGGGTTTGT	1260
GATTGTTCCA ACGGTTAAGC TCGGATGAGG GTCCCTGGAG TCGTAGCTCC CGGGAAACGT	1320
CGACTGGCTT TCCACCTGGA CTTTCATCCGT CCAGGCAGCC CAGAGGGGCT TCAGGCCCCG	1380
CCCGCTCTCC TGCCAACTAC AGCCTCGCGA CTGCGCTCAG CCTTCAGGCC CCGCCCCCTC	1440
GGTCAAGCGG CGTGCTCTCA CTGCACGGCG CCTGGGCCCC GCGCGCCGGG ACCTCGGTTT	1500
CAGCCGTCCT GTCCTGCCCC GAGGCCCTTA GGCCCCGCCC CTGGGCCCCG CGCGCCAGGA	1560
CTTCGGTTTC GACCGTCCTG TCCCGCCCCG AGGCTCCTAG GCCCCGCCCC CTCTGTCCCC	1620

```

GGCGTGTTCT CGCGGCTCCG CCCCTAGGAC CCGCGCGCCG GGA CTTTGGC AAGTTTCAGC 1680
CGTCCGGCCC CGCCCCCTCG GTCCCACGGC TCTCGCGGCC CCTCCCCTAA GTCCCACACG 1740
CCGGGACTTT GGCAAGTTTC AGCCTCCAGC CCCACCCCTA GGTCCCGCCC ACTCGGCCAG 1800
CGGCTGGCTC TCGCGGCCCC GCCCCTGTGC CCTGCGAGTC CCTATTTTGG GAGCATTGCG 1860
GCCGCCGTGC CCCGCCCTC CCCGCGCACC CCGCCCCTCT GCGCGCCCCG CGTCCCAGAC 1920
GCGGGAAGAG CTTGGCCGGT TTCGAGTCGC TGGCCTGCAG CTTCCCTGTG GTTTCCCGAG 1980
GCCTCCTTGC TTCCCGCTCT GCGAGGAGCC TTTCATCCGA AGGCGGGACG ATGCCGGATA 2040
ATCGGCAGCC GAGGAACCGG CAGCCGAGGA TCCGCTCCGG GAACGAGCCT CGTTCCGCGC 2100
CCGCCATGGA ACCGGATGGT CGCGGTGCCT GGGCCCACAG TCGCGCCGCG CTCGACCGCC 2160
TGGAGAAGCT GCTGCGCTGC TCGCGTTGGT AAAGACGGAG CTTCTTGGGG GTGGCTGCGA 2220
GGGCACGGGT CGCACAGTTT CTGGGGGCGG CAGAATCTTT TCAAATCTTC CGTTTCCTCC 2280
TTCCGTTCCC GCGCTGCACT CGGGTCGGCG TCGGGTTAGC ACCTGCCGGG GGATATAGTA 2340
TTAACAACCT CTGCTTCTCA TTCACCTTAT TTTTGGGCGA CTTACCGGCC TCCCCTTGCC 2400
CTGAATCCAA CTGAAACGGT AGTTTTTGAA CTTCAGCGGG CTGAAGAACC GTCTGGAGGT 2460
GTGGCTAAAA AAATGTTTCT CCCGGTCGCG CCTCCAGAGT TTGAATCGGG CTGGGGTGGG 2520
GCTGAGGCTT CTGCATTTTT TACCCGGCCC TGGATTACCC CGCTGCTTTC CGGGAGCTGT 2580
GGCGAATTGG GCTGGCGGGC CGCCCCGGAG ACCCTCTAAA TTAGAAGCAG CTGCCACTCT 2640
AAGTTAACT GGCCTTTTTG ACATTTTCTC CGTGCCAGCT TTTTCGAGTG AGATGGGATG 2700
GAGCATCGGA TATCTACCAT AGTTGTAGAT TGAAGATGGC ACGGAATTTT TCATTTTCTT 2760
AGTTTGCTCA AAAGACTGTA TGTCTGGTGT CCCCCTCTT AGTGATGCTG TTTATTGTTT 2820
TCCTTCATGC TGTCACATTA TGGGAGTCCT CTCAGC 2856

```

## (2) INFORMATION FOR SEQ ID NO: 123:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8804 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 2623..2679
- (D) OTHER INFORMATION: /note= "Exon II"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 5421..6415
- (D) OTHER INFORMATION: /note= "Exon III"

## (ix) FEATURE:

- (A) NAME/KEY: modified base
- (B) LOCATION: one-of(387, 699)
- (D) OTHER INFORMATION: /note= "R = A or G"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: one-of(2743, 5777, 5783)
- (D) OTHER INFORMATION: /note= "W = A or T"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 4763
- (D) OTHER INFORMATION: /note= "Y = C or T"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 6867
- (D) OTHER INFORMATION: /note= "H = A, C or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

```

GAGCTCGAGG ATCAAAACTG TGT TTTTCT GTGTCAAAGG AGAGATTACC TGTTCATGTG      60
TAGATGCAAA TGAGAGAAGA AAGGCAAGTA TTATGAATAT CTAGTGCTTT GGGAGGCAGA      120
AAAGATATGG GAACTATAGG TAGAGTAGCA CATTTTAGGA AGAGTGAATA AAGGAGATCG      180
AATTTTCCTC TTAGACTCTA GGAAGGGTAA GATGACTGGT GAAGAGATAA AAATGTATTT      240
AGGTGTTAAA AAACCTACAC CATTAAGTTC CAGTAAAGTT AATGAGATGA GGAAGCATAG      300
AGATTGTTTT GAATAGCCAT CTATTCATTT GTTGTGTTAT TTATAGAATA TTGAATACAT      360
TACTTTGGTA GAGATACAAA CATGAARAGG CTACCAAATA AATTTTATGT CTTTATTTTA      420
TTTTATTTTA ATATTTTGA GACAGAGTCT GGCTCTGTCA CTCAGGCTGG AGTGCAGGGG      480
TGTAATCTCG GCTCACCTCA ACCTGTGCCT CCTGGGTTCA AGTGATTCTT GTGCCCTCAGC      540
CTCCCCAGTT CCTGGGATTA CAGGCGTGCA CTACCATGCC TGGCTAATTT TTATATTTT      600
AAAATTTTAT TTATTTATTT TTGATACAGG GTCTCGCTCT GTCACCCAGT CTGGAGTGCA      660
GTGGCTCAAT CTTGGCTCAG TGCAACCTCC ACCTTCCARG TTCAAGTGAT TCTTGTGCCT      720
CAGCCTCCCA AGTTGTTGGC ATTACAGACA TGCACCACCA CACCTGGCTA ATTTTGTAT      780
TTTLAGTAGA GACAGGGTTT CACCATGTTG CCCAGGCCAG TCAAGCTCCT GGCCTCAAGT      840
GATCCTCTCA CCTCGGCCTC CCAAAGTGCT AGGATTACAG GCATGAGCCT CAGTACCCGG      900
TTGCTACCAA ATGAATTTAT AGAGAGAAGT TTATTCATCT TTCTTTCCTT TTTTTTTTTT      960
TTTTTCTTTT TTTAATAGTG GTTATTCGCA GCCTGAGTGG GCAGGGAAGA AGTAGACTCT      1020
GGGGCCTATC TAGCATTATA GATTGGCATC CATGAGTGTC TAAGAGGTCA GCACAATTAG      1080
GTAGTGGTGA AGGTGGCTTG GAAATAGTTA ACTCTGGCCT GGGCTGGATA GGAATCCAA      1140
GGTGCTAGGA AACTGAATGG ACTTCTTTCA AGGTAGAGAG CAACCTGAAG GTGGAGGTTT      1200
GGCCAGGGCG ATGTAATAAA AGAGGTAAAG GAAGGATATT ATGATAGGAG GAGCTTGCCA      1260
CGAAATAGAA TTCAGTACAC TTGATGGGGA AAAGGAGGTT AGGTTTGCTA GGTCAAATAA      1320
CAAAAATTAT AGACACAGTA TGAATGTAA AAGAGATCAT GTTTATTGGC AGAGTCACAA      1380
GTCATATTTT TATCTGTACG TTTTACATAC TGTGATACAA AAATCAGACA TGGTGAGTTT      1440
TTAAAAAGTA ATAGGTTTTG TAACGTCAGT GAGCACCATT TTCAGTATCT TGAGAATAAC      1500

```

ATTGTGGGTG	TGTA	CTGTTT	TTCGCAATCT	TTCTGAAAAT	CTCAATTTCA	CTTGAATGTT	1560
CATATCTATT	AAGATGTGCA	GTGTGATACT	TTCAATCTTT	CAAGAAATGT	GAAGAATTGT		1620
TTTTATTGAT	ACGTGGTATG	TGTACAGAGG	TAATTTAATA	ATAATGGTGT	TAAATATCTA		1680
AAGGTTTTAG	AGTTAGTATA	ATAAACCAAG	CCAGAAAAAG	TGCTCATCAT	TTAAAAGGCT		1740
TTACTTCTCT	GGGTACATTA	CATCCATTGA	GTATAATGTC	TTGGTGTGTA	TTTATTAGTA		1800
TCATTACTTT	GTTAGGATTA	ACAAATGTAG	CAGGTATTTT	GATGGACAGT	ATTAGAAATA		1860
TTTTGTTTCT	GTGTCTCTTT	GCTCACTCAC	TGCAGGGAAC	TTCAATCTAT	CTGCAGACAC		1920
ACTGCTAATC	TGATGATTTT	TGTTTACATT	TGCTTAAAGA	ACAATTAACT	TTCTGTGAAC		1980
TGAAAGAATG	GCTGCATTTT	TCACAATATA	TCTGTGAAAA	ATTGATGGCT	ACTTTACAGT		2040
AGTAAGAAAC	ACATGTTTTT	TTTAATTGGG	AAACCTTGA	TTCTTACGGT	CAAAATGACT		2100
AGATATCCTG	GTTTGTCTGG	GGCAGTCCTG	GTTTATTCCT	GTTGTCCCAG	CATCCTATTC		2160
AATTATCACT	CCTCTTCACT	CTCAGAAGTT	TTCTCATTTG	GATGATACGT	TATATGGTCA		2220
TTCTACGTGT	GGGACACATT	TTGAGGTATA	TGGTCTACAC	TTTGAATCAT	AAGGGGAAGA		2280
TACACCTTAG	TAGTTGAAAG	AAGCAGCCAG	TAGTTGGTTT	TGCCTTAAGA	GGTTTAGAGA		2340
TACTTGAACA	AAATTGTTCC	TTTCTCCAAC	TTTCTAGAAA	ATACATTTTA	AAATGTATTA		2400
CAACTTGTA	CCAGTTTGGT	GGTTACTATT	TAAAATCATC	AGGTATGTTA	TGGTACAATA		2460
TTTAACCAGG	GAGTAACAGC	CTTTCAAATG	AATGCATCCT	TAATACCTTC	TGCTTTGAGA		2520
AAGTGAGAAA	TATGGTAGTG	TTGGGCCTTG	GATGAAATAT	TGGGTGTGAG	ATGTTTATCT		2580
AACAATGACT	CAAATCTTGA	TTGTTTTAAT	TTCTTCAAAC	AGTACTAACA	TTCTGAGAGA		2640
GCCTGTGTGT	TTAGGAGGAT	GTGAGCACAT	CTTCTGTAGG	TAAGTAATTA	CGGTTTGATG		2700
TATATAGTAC	AAGTGTATTT	TTTACTAGAT	AACAATCATG	TAWTCTTGTT	GATAATAATG		2760
GTACTTGATG	TTTGTGTAAC	TTTCAAAGTC	TGCAAAGTAA	CCTATTGTAC	ATTATCTTAG		2820
TTTGATCTTT	AAAAGTGTGT	CTTAATTTAC	AAATAAGAAA	ACCAAAGCTC	AAAGAGGATG		2880
ACTTGTCCTA	GTTACAAAGC	TTAGTAGACA	GCTTTGCCAA	ATTGGAGGAA	AAAAATTAAT		2940
GCCTTTTATA	TAATTCAAAT	AGATGTTTTT	AAATTTTCCA	GTTAAATTTT	GAATCTAGAC		3000
TCAAATTTGT	GAAAGTATAG	GTCTTACAGT	TATTATATTA	GTTTCCTAAG	GCTGCCATAG		3060
CAAAGAACCA	AAAAGTGGGT	GGCTTAGAAC	AACAGAAATG	TATTGTCTCT	CTAGTTCTGT		3120
AGGCTAGGAG	TCTAAGATCA	AGATGATGAC	AGGGTTGGTT	CCTTCTGTGG	GCTGTGAGGG		3180
AGAATCTGCT	CCTTGCCTTT	CCCCTAGCAA	TCTTTGCTGT	TCCTTGACTT	GGAGATGTAT		3240
CACTCTGGTC	TTTGCTTTCA	TGTACACCTG	GCATTCTTCC	TTGAGTGCTT	GCTTCTCTCT		3300
GTGTCAAATT	TCCTCTTTTC	ATACGAATAC	CACCTATATT	AAGGGCTCAC	CCTAATGTTC		3360
TCATCTTAAC	TTGATCGTCT	AGAGCCTCTT	TTTCCAAATT	AGGTCACATT	CACAGTAACT		3420
GAGGGGTGAA	GACTCAAACA	TCTTTTTGGG	GGACACAATT	CAGTGCATAA	CAGTTATTGT		3480

GAAATTATAT CCATGTGATG GCCCTGGCTT ACAGGTCAGA AGATTAGATT TTTATCTCTT	3540
ATTTCTTGTT CAGGAGAAAT CAGTTGAGAG ATTAAGTGTC TTAGTTTAAT TGCCTATGAG	3600
ACAGGGAAAA TATAGTCTCC TTTGAATTAA TCTTTTAATT ATTTCCAGTT ATGAGTATGT	3660
TACTGTCTGA CATGAACAAA TACATTCGGA AATTTGAGCA TATATGAAAA TAACCTTGTTG	3720
TATTACTTCA AGGGCTAAGG TTGTCTGGGA GGCAGCTATG TTTTGGGTCC CAATTTTGTT	3780
CCTGGAAAAA CAGATCTGTT ATACGAGGGA TATCCAGCAG GAGGATGGGG GGTAGGAGAT	3840
GGAACCTGTT GCCTACCAGA TACGTCTGTT TGGATGATGA AAGAAAAGGC ATTTTAAAGT	3900
TGGTGTGTTT AGAGCTGACC TCTTGATCAT CCTCCTCTTC AGCCTTTTCC TGCCTGGGTA	3960
ATCTTCATCA CAGTAAATGG TACCAGTTAC TGCTAGGTTG CTTATCCCAC ATCGCTAAGT	4020
TCTGTCAGTT CTGCCTCAAA ATGTGTCCTG AATCTTGAAT TTTTCCACCT TTTTCTGAA	4080
TTTTCTACAG TTTATAGTTT GTTCCAAGCC ATCATTGTAT TATCTCTGCT CATACTATTG	4140
TGATAGGTTT ATAGCTGGTC TTCCCTTTTT GGCATCATCA CCCCTTCCTC ATTTTCTACA	4200
TTCTTGACAG TGATCTTTGA AAAATCACGA TAATATCCAT ATCATTACCC TGCTTAAACT	4260
CTTTAGTGGG TTCCTATTGC AGTTAAAATA AAATCCAAGC TCTGCCCTCT GGTCTGCAAA	4320
ACCCTGTATG GGACCTAGAA CCTGTCTTCC TCTTGGACGT CATTTACCTC TGGCTCACAC	4380
TGTTCCAGCA CTTCTCCTTT TAGCTCATTC AATACACCAA ATTCACTCCC GCCCCAGGAC	4440
TTTGGCACTT GCTATTTCTT CCACGTGAAA TGTACTTATG CCAGATTTCT GTGAGGCTTA	4500
CTTTTTACCA GTTATATATC AGCTGAAATG TCACTGCCTC CAAAACGCCT CCCATGGATC	4560
TGCTTACGAA ATGATGCCCA CTCCCTGTCC CCACTCTGTA GCAAGCAGAT GTATTTTGTT	4620
TCAGCTGACA GATGATGGGT TTGCTCACTG TATACAATAG GTCAGTCACA AAGCTGCTAG	4680
AGCATAATGA ACTCCATGAC ATTTTATGCA AAGTTGGTTT GTGTGTGTGG TGGGGTGGGG	4740
GATCCATACG TTTAATCAGA TTYTAAAGGA CTCTGTGACT TAAACAACC AATCAAAGGT	4800
TTGAGACACG AGAACATCTT AAAAGAGAAA ATGATTATGT GTATATGACA TGAACGTAGG	4860
AAAGAATTCA TTCAGTGCAG TTTTGATTG TTTTGATTAT ATGATGCCAA AAATTATGGT	4920
AGCTGTTTTT ATTCATGCAG ATTTTGAGTT AAAACTGCTC AGCAATGAGG TTTTAAATG	4980
ATTTGACATA GCTCAGTTCA ACTGATAAAG GTAATTCATC TACTCTCTAA GATACAATTT	5040
AAGACATGTG GCAGGGGTGC TCAAACATTT TGAGCATTCT TCAGAATTGA GAACCATAAA	5100
GAGCTTTTGT GTGTTTCATGC ATTTACCGTA CAAAAGTGA ACTTTTAAA AAAAGAATTA	5160
ATTTAAAAAT AATAAACCCA TTAATACAAA CACAAATAAC ACTTTGTAAG GAAAAATAAC	5220
AATTTTAAAC AATCAAAGAA AAAAATAGAG TGGCACTGTT GTACATTTT GCAAACTCTT	5280
GTCTTGCTGA ATGGAAAACA GCTACATTCT CAACTTATAT ATTCAGTCTA TTGTGATACT	5340
TTTTTGATTG AAGAAAATCT GGCTTTATAC ATAATTGTAG TTGGAAGAAG TATTGAAGTA	5400
TTTTCAGTAG CTTTTTTAGA TAATTGTGGC TATTCTTCTT TAATACTACA CAAAAGTGTG	5460

ACAAGTGCTT	TTTTATTTTT	ATTTTTTTGT	GAAGACAAGA	GTCTCACTCT	GTCACCCAGG	5520
CTGGAGTGCA	GTGGCGCGAT	CATGGCTCAC	TACAGCCTCA	ACTTCTTGGC	TCGGGTGATC	5580
ATCCAATCTC	ACTTTCTGAG	TAGCTGAGAC	TACAGGTGCG	CCCCACCACG	CCTGGCTAAT	5640
TTTTGTAGTT	TTTTGTAGAG	ACAAGGTTTT	GCCATGTTGC	CCATGATGGT	CTTGAACCTC	5700
TGGGCTCAAG	CGATTCTCTT	GTCTCGGCCT	CCCAAAGTGC	TGAGATTACA	GGAATGAATC	5760
ACCATGCCTG	GCCAACWAGT	GCWGTTTTCT	TTAGGTTTCAT	CACAGTGTGG	AATCTGAAAC	5820
TATATCAATA	AATTTTTATA	CTCTGTTACA	TGAAATTTCA	TTGGTTTATC	TAGTACTTTG	5880
AATTTATCTG	TACTCAGCA	TGATTTTATG	ACATCACACA	CGGGTCATTT	GAGAAATACT	5940
CACTGAGCTA	TACAGGTCCA	CCGAAAAATG	ACATTTTTTT	CAGAATTCCA	AATTAATGTA	6000
TATAAAATAA	TTTGGGAATA	GTTTATAAAA	TGTTATAGAA	ATATAAAATA	CTAGTCTGAT	6060
CTGAGTACTA	GTGCAGTGCT	GGACAAGTAA	CTTAAATAAC	TGGATCTCAT	TTTCTTCATC	6120
CAAAAATGAG	GTGGTAGTTC	TTGATGAGTT	GATTTCAGAT	TGTATTTTAC	TGATAATTAA	6180
GTATTGCACA	AATGATTTAA	ATTGCATGAG	TAATCAGTTT	TACATATTTT	TTTGTGTTGG	6240
GGTTCCAAGT	TAGAGTTCTT	AACTACTAGC	AAACAAATGT	ACAGAAGATC	CTTTGCTAAA	6300
GAAAGTTGAC	ATTTATTGAC	CCCAGTGACA	TTTTTTGAAT	TAGATGGTCC	CAAAAGTCTC	6360
TTCCAGCTCT	GGTATGGGTT	ACAGATTCAT	TTTACAATTT	TTTTTGAGTT	ACATTCTGTC	6420
AGAAATCATG	CTAGAAGCCA	AGGATACAAG	AATTAGAAAT	GGCATAGGTT	TTGTTTGAGG	6480
ATAGTTCATA	TTGAGTAAGA	ATCCTCTCTG	CCTACAGAGG	ATTGGGTCCT	GTGACAAGGA	6540
ATGTCCTGTG	GTGCTTGGGG	AAGATGTGGC	TTTTCAACTG	TTACATTACT	TACTCAGTCC	6600
TACTGGAACC	CATCTGGTGA	GGCCAATGAA	GGAGGAGATT	TATAGAATTC	TTATTCTGGA	6660
ATTCACAGAT	GGCCTTGTGG	GGGGGATCGT	GAATCCCTCT	TTTTACATGA	GTATGTATAG	6720
ATTTTCATCA	GATTTGCAAA	AGGTCTATAA	ACCCCAAAAT	TAGAAAAACT	CTTCTAACTC	6780
TGAAACTTTA	GTCCTTAGAT	AACCCAGGTA	ATTGAGTCTA	CAGGTTTTAA	AATTGTCTGA	6840
AAAAGTCAAA	GATCTTTTCC	CAAAAGHTAC	TTTAAAGCCT	TGAGATACTA	GTCCCAAAAC	6900
AAAACAAAAG	GTCTTTCCCA	GTGTCTGCAG	TAGTTTTGGG	TATTTTCATG	CAATGTTAAG	6960
ATAGAAAAAG	TTAGGATGCA	CGACTACTAT	GCATTAGGCA	TTCTATCTTT	CTGTTACATC	7020
TCTGTAACCT	TTAGAAGCTA	CAGATTTATT	TGTAGGGAAA	ATTATGCAGA	CTAATAATCA	7080
GGCTAAGTAA	AGTCTCTTCA	TAGCAAATTA	CATGAGCAAC	CTTAGATTTG	ATGTATGTAT	7140
TTTACTCTTT	AAAACAGTAT	TCAACAAGGA	TATTACAATT	GACCATTGTA	TGTTAGAATA	7200
ACCTCTGCTC	CATTTATTTT	TGTTCAAAC	GTTTAGTTTT	TGGAATTAAA	TTCTGCTGAA	7260
TGGGTTGCTT	TTTTTTTTTT	TTTTAATTAT	TTTAAAGTAA	TTGTGTAAGT	GACTGCATTG	7320
GAACTGGATG	TCCAGTGTGT	TACACCCCGG	CCTGGATACA	AGACTTGAAG	ATAAATAGAC	7380
AACTGGACAG	CATGATTCAA	CTTTGTAGTA	AGCTTCGAAA	TTTGCTACAT	GACAATGAGC	7440

TGTCAGGTAA GAACTATCCC TATCTCTCTT AGTTAAATTC ATCAGTTAAA AACTGATGAA	7500
TTCATATTCA TAAAGTATAT AAAACATCTA TCTGGAGTTC TGGAATACGT ATTTTCAGATT	7560
TTAAAATCTG TAGGTTTTTT TTTTTTTTTT TTAAATAGCC ATTGAGTCTC TCTATGTTGT	7620
CCAAGCCGGA CTTGAACTCC TGCACCTCAAG GGATTCCCCC CACCTCAGCC TCCCTGGTAC	7680
ATGCATGTGA CACACCCGGT GGTTTTTTGG AAGATCATTT TGAGAAATGA CGGGAACCAT	7740
AGAGATAATT AAAGTCTAAC CCCCTTATTT TATAGTTGAG AAATCAAGGT CCAGAGAGGT	7800
GAAATAACCT TGCCTACAGC CGCATGTGTA GTTAGTACTA GAAGCTGGAG AAATAAAGCC	7860
CATTTCTGAC TGTAATCCA GGAACTTTTT TGCTGCTTTA CATTGCCTGC TTTTACATT	7920
TATGATAACT TCTGCAGAAT ATATATGGAA TAGTGATTTT GGCCTTAATA GCACTTAACT	7980
CACTTAAATA CACTTTCCCA AGATAGAATA CGACATTTTT CCATGGATCA CTTTCTAGA	8040
CTGAATTAAT TAGATGAGCA CATTTTGAAA GAGCAGAAAT CTAATATTCA TTTTCTTTTC	8100
TATTTAAGTG GGGGTAAACG TTTTTTAAAT TGTCCTTAGA CTCTGTGTAT TAACTGTGTA	8160
TCTTTCCATT GTGCTTCTG CTGAGAACTA ATTAAGTGTG GAAAGAGAGT AAAATGTTTT	8220
GTATCTTCAT AATTGGATAT TAGAGTTGTC TTTTATTGA CCAGATCATG CTATTTTAGT	8280
GTGTGTTTGT AGAAGAACCT GTTCTTGACT GGCAGGATGC CATGGATGAT TGATAATGCT	8340
CATATAAAAT TGTTAGATTT CTATTTTTAA ATCTTTTGT CTTAGAATCC TGCCATGATG	8400
TCTTATTGTC AGCTAAAGAT GAAGTTGATT ATACAAAATA AATAAATGTG GCAAAAACCT	8460
TGAGTCTTAC CACAGGTCCA TATTTTAAAG AATTAGAACT TAAATCACTT TATACCTATT	8520
AAAATTTTCC TTCAGCTAGT TATCTGCGTC ATTGTATTCA ACTCCATTCT TTTATGATAA	8580
AATGTGCTGT AGTGCAGAAG TTTTCTTAC TTTCAGATGT AACTATACAC ACACATTTTT	8640
AAAAGTTGCC GTTTTTTAAA AATGATATTG TAGTTGTAAA CCTTTTAAAG AACACACTGA	8700
AGAAAAATTC TGTCATTAGT TCATCTGAAC TCTTCTTATT TAAAGACAGT ACTGACAATG	8760
TTTATTTTCT AGTTAAAAGA GATTGGTGTG TGATCCTCGA GCTC	8804

## (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2111 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1899
- (D) OTHER INFORMATION: /note= "D = A, G or T"

## (ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 621..1570
- (D) OTHER INFORMATION: /note= "Exon IV"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

CCCTTTATTC ATTTTGTGAG AGATACGTTT TGAAGTCATA TAGTATGAGT CCTCCAACCTT	60
GGTTCTTCAG TATTATGTTG GTTGTCTCTAA GTCCTTTGCC TTTCAAATTG TACCTTTCTT	120
TTTCCGTGTA AATTTTATAA TTGCTTGTCG AGTTTACAA ACAAGCCTAC TGGAAATTAC	180
ATTGGGATTT GTTATATCTG TAAATCTGTT TGGAAACAATT GGCATCTTAA CAATATTGTT	240
ATAGTTCATG AACATGGTAT ATCTCTCCAT TTGTGTAGAA TTCTTTAATA AGTATTTTGT	300
AATTCTTAGC AAACAGAACT TATACATTCT GTTAGATTG TATTTTATGG GTTTTTTTGG	360
GTACTATATT TGGAAATATAA GTTTCATGTT TTGTCTCTG CTATATCTTC AGTCTCTAGG	420
ATATGGCACA AAGTCGGATT CAGTAAGTAT TGAATGAGTG AATATCTGCT ATCAAAGAGT	480
TCACACTCTA GGAGCTGAGA AAGAAGTACA TAATTAAAAG ATGATACACT TTAGGGGAAC	540
TGTAAACAAA ATTCTTCGGG AGCTCCATGT GGGAGCAATA AATTCATGT AACAGATTC	600
TTTTTCTTTT TTTCTGTCAG ATTTGAAAGA AGATAAACCT AGGAAAAGTT TGTTTAATGA	660
TGCAGGAAAC AAGAAGAATT CAATTAAAT GTGGTTTAGC CCTCGAAGTA AGAAAGTCAG	720
ATATGTTGTG AGTAAAGCTT CAGTGCAAAC CCAGCCTGCA ATAAAAAAG ATGCAAGTGC	780
TCAGCAAGAC TCATATGAAT TTGTTTCCCC AAGTCCTCCT GCAGATGTTT CTGAGAGGGC	840
TAAAAAGGCT TCTGCAAGAT CTGGAAAAA GCAAAAAAAG AAAACTTTAG CTGAAATCAA	900
CCAAAAATGG AATTTAGAGG CAGAAAAAGA AGATGGTGAA TTTGACTCCA AAGAGGAATC	960
TAAGCAAAAG CTGGTATCCT TCTGTAGCCA ACCATCTGTT ATCTCCAGTC CTCAGATAAA	1020
TGGTGAAATA GACTTACTAG CAAGTGGCTC CTTGACAGAA TCTGAATGTT TTGGAAGTTT	1080
AACTGAAGTC TCTTTACCAT TGGCTGAGCA AATAGAGTCT CCAGACACTA AGAGCAGGAA	1140
TGAAGTAGTG ACTCCTGAGA AGGTCTGCAA AAATTATCTT ACATCTAAGA AATCTTTGCC	1200
ATTAGAAAAT AATGGAAAAC GTGGCCATCA CAATAGACTT TCCAGTCCCA TTTCTAAGAG	1260
ATGTAGAACC AGCATTCTGA GCACCAGTGG AGATTTTGTT AAGCAAACGG TGCCCTCAGA	1320
AAATATACCA TTGCCTGAAT GTTCTTCACC ACCTTCATGC AAACGTAAAG TTGGTGGTAC	1380
ATCAGGGAGC AAAAACAGTA ACATGTCCGA TGAATTCATT AGTCTTTCAC CAGGTACACC	1440
ACCTTCTACA TTAAGTAGTT CAAGTTACAG GCGAGTGATG TCTAGTCCCT CAGCAATGAA	1500
GCTGTTGCCC AATATGGCTG TGAAAAGAAA TCATAGAGGA GAGACTTTGC TCCATATTGC	1560
TTCTATTAAG GTAGGATGCT TACTCTGAAA TACCATCTCA GAATGAGGCC AACTATAAAG	1620
CAATTTCTTT GCAGTTTTTG AAAAATGGCA TAGGATTACT AGGATAATTA ACCTTTCACA	1680
GACATGATAC TTCCTCTGAA CCAGAGAAGC CAGATTCACA GGGAGAGCAT CTCTACTTCA	1740
GTTGGAGCAG TGGCCCCTGA GTCTGGGCGC ATGATCTTGT AGGAGAAAAC CAATATTTGA	1800
ATATTTTACG TTTTATTTTG CCAAGTGCTT TTGCTTTTGT CTATTTTACC TTCAGTTTTT	1860
ATCATTTTGT TTACCTGTCT TCATGCTTTA TGAATGTADA CAATTGCTAA GTTATTACAG	1920



GCAACAATGT TTA CTTAGTA AAAAGCCCCA TATTTACCAT CCAAATTCAA CCAAATTTG 1980  
 GAAGGTTGAA AGATGTGGTC TGTACATTTT TCCAATGACC GGGACATTTG ACTATCAGAA 2040  
 ATGGCTCCTC CAGTTCACCA CAAAGGAGCT GCTTTTTACC CTACAATCAG CTGTTCTTTT 2100  
 TACTGACCTG T 2111

## (2) INFORMATION FOR SEQ ID NO: 125:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1098 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 451..531
- (D) OTHER INFORMATION: /note= "Exon V"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

TTCTTTTTTG TTTTGT TTTT TTGAGACGGA GTCTTGCTCT GTGGCCCAGG CTGGAGTGCA 60  
 GTGACATGAT CTTGGTTCAC TGCAACCTCT GCCTCCTGGG CTCAAGTGAT TCTCCTGCCT 120  
 CAGCCTCCCA AGTAGCTGGG ATTACAGGCT GGCACCACCA TGCCCGACTA GTTTTGTATT 180  
 TTTAGTAGAG ACGGGGTTTC ACCATGTTGA CCAGGCTGGT CTCAAACCTCC TGACCTCAGG 240  
 TGATCCACCC GCCTCGGCCT CTCAAAGTGT TGGGATTACA GGCGTGAGCC ACCACACCCG 300  
 GCCTAATAAT TTATTAATC ATGAACAGTA GCCTTAAGAG AAAACGATTT AAGTTTTACT 360  
 TTATATTGAA GAAGGCAGCA TTTAAAAAAG CTCAATATTT TCCTTTCTTT CCTTAATGCT 420  
 TTTTAATTTT CATTTTGTTC ATTTTCTAG GGCGACATAC CTTCTGTTGA ATACCTTTTA 480  
 CAAAATGGAA GTGATCCAAA TGTAAAGAC CATGCTGGAT GGACACCATT GGTAGTTGTC 540  
 TGGTTTTTAT TCTCATTCTT TCTGTGTTTT ACAGTTCTTA TAGTTTATAG TTATGTAGTT 600  
 GTCTATATAT CATCCTCTGC CACATATACT CTTTTTAGTC TGAAGAACTT ATGTTTTCAT 660  
 CAAGTATGAG AACATGATTA CTTTCCTTCT AGCTTTTCAT TTGTGACAGG CAAGAAATTG 720  
 GTTACCTTTT GACAGACTAC CTTTAGATTT AGGAATCCAT TTGTACTGTA CTGCAGAATT 780  
 TAGCTAATGT CTAGAGGTAA CAGCTACAGC TGACATCAGG CTCCATTCTG TAGCACTGCA 840  
 TGTCACCTGA ACCAAATTTT TTGGAACAAA AAGAGGTCGG AGGAACTGAG TATAGGAAAG 900  
 TGATCACAAG GAAGTAATTC TCACTGAGGG TCTATCTTAG CCTCACTTAT ACCCTATCCA 960  
 ATTGTAGATA TATAAGGCAG TAGAAATCTT TGCTTACATT GAACATTTTT AAAGGTCTTT 1020  
 GCTCATTATT ACTAAAAAG TGTGAAGCAT AATCTGGAAA CAGAATGACA CAAATGCTTG 1080  
 GAAACAATTG GTATGTAG 1098

## (2) INFORMATION FOR SEQ ID NO: 126:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1756 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 508..680
- (D) OTHER INFORMATION: /note= "Exon VI"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

TTGTATTCCT ACTCGTGTTA TTTCTCTTAT CAATAACAGC AGCCATACAG CAATTTGTAG	60
GTTTCAGCAAA TAGCTTGTCT TAAAAGCATT CCTTCACGGA TACTTACTTT GTTGCATGAT	120
ATGTGTATGT ACTGGTACAG ATTTTATGTA TCATCTTGTT ATTAAATATG TAGACTTTTT	180
TTCTTAATGT GTTACATTTA TTGTAGAACA TTTAAGGAGC TACCGTAGGT TTAAAACTAC	240
ATTTTCTTCT AAAAAAAGA AAAGTGCTTG ACCCAAGGCT CAAATGAGAA TAGCCTTTCT	300
TTTTTTATGA GTTACACAGA TCTTGATTGA AAGATTATTA ATAGTAACTT TCACTCTGTC	360
AGCAACTTAT AGTGTTTTTG AGTATTTAGG TAACAATAAA TTTACTGCCT GACGTTTACA	420
TTTATTTTTT TAAAGTGTA TATTATAATA TCATCCATTG CTCTTTCTTA TCACTTCTTT	480
CACTTCTTTT TCAAAAAATT TAATTAGCAT GAAGCTTGCA ATCATGGGCA CCTGAAGGTA	540
GTGGAATTAT TGCTCCAGCA TAAGGCATTG GTGAACACCA CCGGGTATCA AAATGACTCA	600
CCACTTCACG ATGCAGCCAA GAATGGGCAC GTGGATATAG TCAAGCTGTT ACTTTCCTAT	660
GGAGCCTCCA GAAATGCTGT GTAAGTAGTT CAATGTAAAA ATTATTTTTA AAATGGACCT	720
ATATTCTTGA GTCAAGGTGT GTGATAAAGC AGACTTTAAT AGTCAAGTTG ATGGCTTTCT	780
TCACTTTCAC AACTAAAATT AGATGTGATC ATCACATTCT GCACTCATAA TCAGCATTCA	840
TGCCCTTTCT CTTTATGATA CAGTTGGTCC TTCATATTCT TGGGTTCTAC ACTTGAGGAT	900
CCAGCCAACT GCAGATCAAA AATAATTGGG AAATATCAAT GACAGATCGG ATAAAGAAAA	960
TGTGTTACAT ATATACCATG GAATACTATG CAACTACAAA AAAGAATGAG ATCATGTTTT	1020
TTTGTGGGCA CATGATGGAG CTGGAGGCCA TTATCCTTAG TAACTAACG CATGAACAGA	1080
AAACCAAATA CCGCATGTTT TCACTTATAA GTGAGAGCTA AATGATGAGA ATTCATGAAC	1140
ACAAAGAAGG GAACAACAGA CACCAGAGTC TACTTGAGTG TGGAGGGTGG GAGGAGGGAG	1200
AGGAGCAGAA AAAGTAACTA TTAGGTACTA GGCTCTATAC CTGCGTGGTG AAATAATCTG	1260
TACAACACAC CCCCCTTACA CAAGTTTACC TATATAACAA ACCTTCACAA CTAAATAAAA	1320
AACCTAGAAT AAAAGTTTAA AAAGGGAAAA AAAAATAACA CTACGATAAT AAGTAATATA	1380
GGTAAAACAA TATAGTATAA ATATTTATAC AGCATTTTCA TCTATTAGGT ATTACAAGTA	1440
ATCTGGAGAT GATTTAAAGT ATACGGGAGG ATGTATGTAG TTTACAAGTA AATACTATGC	1500
CATCTTTTAT AAGAACTTG AGCAGTGGCA CATTTTGACA TCACAGGGGT TGAGGAACCA	1560

ATTCCCCATG GATAGCAATG GGGATAATTG TGCTGACATA TTTGGGGGAG ATTTACTTTC 1620  
 TTAATTCAGA AACAGTTGTC AATTTTGGAA GCTTTCATTT AATGGAAAAA TTTACTTAGT 1680  
 GTTTATATTC TGTAGATTGA TTTACACTTT AATAAGCAGT TATTGTAGAA ATAATTATTT 1740  
 TGTATGCTTC CTAATA 1756

## (2) INFORMATION FOR SEQ ID NO: 127:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1190 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 548..656
- (D) OTHER INFORMATION: /note= "Exon VII"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1023
- (D) OTHER INFORMATION: /note= "W = A or T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

TGAGCCATAA TACTTTGTTC TGCGATGGTT GTGATTATTA TAGGTTATTG TATGCACACA 60  
 GTTTAAATT AATTTTAAA GTACCCTGTT AACTATATTA TTAACTGTT TGTTATGTGG 120  
 CATAATTTTC CTTCTAGTAG AACAAATCC CTGTCCTGTG AATTTATCTA ATTTTTTATT 180  
 GGTTTATAAA GACTATATGG CCTATAATAG CTATAGTAAA TGATTTTTAT TGGCATTGTA 240  
 AAGTCTGTCA CTTATAGTGA TTGGTGATTA TGAAGCCATA TTTTAATATG AATAAGAATG 300  
 CAGAATACAG TTGTGAAAAA TTCATAATAC TATATTCAGT AAAACAATC CCTATAATCT 360  
 GATGTCAAAC TGAAATTTTA CATCATTTCT CCTTTGAGTT CAGCAGCTTT TGATTCTAGA 420  
 TTCTTCTGCC TAATATGAGT TCTGAGTAAT TTATTTTAGT TAAAATTGTA TATTATTAAG 480  
 GATGTTGAAA AATTGAGTCG AGTCACACAT TTGACTTACT TAAACACATC TGCATTATT 540  
 TTACCAGTAA TATATTTGGT CTGCGGCCCTG TCGATTATAC AGATGATGAA AGTATGAAAT 600  
 CGCTATTGCT GCTACCAGAG AAGAATGAAT CATCCTCAGC TAGCCACTGC TCAGTAGTAA 660  
 GTATGGATTT AGCTTTGGGA CATTTATATA TTTTATTAAA ATTGGTTATG AAAGGAACAT 720  
 AATAGAAAAA TTTCCATTTG ACCAATTGCT TACATTCACC AAACAATTAT TGAGCACTTC 780  
 CTGAGTATTA GCTACTGTGG ATTCAAAGAC ATAATCACAG TACGACCATC TAGAAATACT 840  
 TATTGAGCCC ACTCTGTATT TTAGGCAGCA TTCATAAAAC AATGAATATG ACTGGTAGAA 900  
 CTCTTATTCT CAGGGAGGAG CTTACCATCT GAGAGGTAGG AAAGAGACAA ACTGTAAATA 960  
 TTGAACAAAT ATAAATAAAA TAATTTTCTA CACTTAGACA TGAGTGTTTC GAAGATGTTG 1020  
 TAWAGTGTCG TTGGGTGGAG GTAGTGGGCT TCTGCAAGGC CATTGCTTTA GCTAGGGTCA 1080

CAGAGTGGAG CCTTAAAGCA CTGATTTGAA TTGAAACCTG AATGTTGAAG TGAGGAGGCT 1140  
GCCAGGTGAC TATCTGGAGG ACACAGTGTA TAGGCCCTTC AGTGAATGAG 1190

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 915 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 566..698
- (D) OTHER INFORMATION: /note= "Exon VIII"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

CTTACTATTT ATGGATCTGA TCTCCTAAGT TTTGAATTAA CTTGTCTGTT TTTATCTTTT 60  
CCTAGTTTTG AGGGGTTACT ATTTTGATGC TAATTTGTTT TCTATCTTTG AGGTCAGCAC 120  
TGTTCTAGAA GCCTTGGCAT TCTTTGATTT TTCAGATAAT CTCAGTTTAA ACTAAACAAG 180  
TTTGATTTTA ACTCTATTGG GACAAGTTAG TGGAGGTGGA ATAGGGAATT GCTGATTTTA 240  
AGTGGATATT TTAAGTTACT TGGGAAAAGA AAAAGACTTA CTGGTGA CTG AATGAAGTAA 300  
AACCCTAGAG AGACCCAATT TAAAATTGAA GAAATGAGAT GCCCCTGGGT ATAGAGAGCT 360  
ATCACAATTG ACATTTTCTT GAGGGAAAAA TAAAGAGAAA AAAATTATTT AAAAGGTTCT 420  
GGGTGTAGAT TCAATGGAAA TAATTGAAAA TTATTAGAGT AAATAAGTA ATGAAATTCA 480  
AGCTTATATC AAGTAACAGT CTGTTTAATG TCTTTGTCTA GTCGTCTAAT GTTTTTAACA 540  
CTGGTATCTC CTTTTATATT AACAGATGAA CACTGGGCAG CGTAGGGATG GACCTCTTGT 600  
ACTTATAGGC AGTGGGCTGT CTTCAGAACA ACAGAAAATG CTCAGTGAGC TTGCAGTAAT 660  
TCTTAAGGCT AAAAAATATA CTGAGTTTGA CAGTACAGGT GAGGATTTTG AATTTTGGGA 720  
GGTGGGGTAG AAAAAATGTT AAATAGATGA TCCTTTTGGG GAACTACCTT TGATAATTTA 780  
CATATGTTTT AACCATTGGG AGATGGCTGT ATACTTTGCA TCTTGTAATA AATCTAAATT 840  
TTTTTTCAGT AATAAACTAC TTATAGACAA CAACGTAGTT AGGAAATGTA AAGTTTAAAG 900  
GTTTGCATAT ATTTT 915

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 464 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 226..318
- (D) OTHER INFORMATION: /note= "Exon IX"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

```

CAATATGGCT TTAAGATATA TGGTTTATGA TCTGATTTTT TATATTGATG GCCAGGTTAG      60
AGAACTAGAT ACTAAATAGA AGTAGTCTTA CACTTAAGTG TAAAAATTGT TGCCTTTGAA      120
GATTCAGATA TAAGCTTACA AAATATAGAT GAGTTATAAG AAGCAGGCCA AAGAAATACT      180
TTGGCTTGTA TCTTTCTTTC TCTTACTGCT TTTTTTGAT TTTAGTAACT CATGTTGTTG      240
TTCCTGGTGA TGCAGTTCAA AGTACCTTGA AGTGTATGCT TGGGATTCTC AATGGATGCT      300
GGATTCTAAA ATTTGAATGT AAGTGTGGA TTTGAGAGAA TTAAGAAATG AATTAGACTA      360
GTTTTGTTTT TCATGGTTAT TAATGCCTGT GATTAAGGAA CTTGATGTTA ATTTTCTTAC      420
CTCTGGTTAG TCACTGCATT TTGGAAAAGC TTCTGGCTGG GCGC                        464

```

## (2) INFORMATION FOR SEQ ID NO: 130:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4334 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 519..616
- (D) OTHER INFORMATION: /note= "Exon X"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 2019..2351
- (D) OTHER INFORMATION: /note= "Exon XI"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

```

CCCTGTTGTG TGGCTAGCTG AGCTTGGTGC TGTAGACTAA AGCACATTCC TTCATGTCAA      60
ATCACTTACA GTTTAACAGA CGATTAGACA TATAACTGTC AAAATAAGCA GTATAGATGG      120
TAAGTGCTCA GTTTAGGTTA TTGTGTCATG GACTTTTTAT TCACCTTAAT TTTGGGTAAT      180
TGCTATGAGT GGAAATGTAG ACTTTTATTT TTGTCTTTGA AATAGTATCC TGGCTTAGAT      240
TTTTTCAGAAA GGAGATTAAA ATTACAGTTA GTGTTCACTA CTAAGTTATG GCTTAATCCT      300
CCAAATAAAG AGTTTTTTTAA AATATTTTCT TTATATGGGA AAACCAAGTTG TATTACATTT      360
TGTTTTGGCA TAAGTAAGAT TTCTGTTTGC ATTTTAGAAT AATACTTAAA AACTGCCATG      420
AAGAAGAAAA ACCACTTAGG TAAATTGCTT GATTTTAAATG AGAGAGATAT AGTGCTCACT      480
TGATACTTAG TTTGCTTTAA TTCTTGTTGT TTTGTCAGGG GTAAAAGCAT GTCTACGAAG      540
AAAAGTATGT GAACAGGAAG AAAAGTATGA AATTCCTGAA GGTCCACGCA GAAGCAGGCT      600
CAACAGAGAA CAGCTGGTAT TTTTCTTTTA ATACAACCTT CATTGTTCTT ATTATGACAT      660
ACTATTATTA TCACCATCAG GAAGAACTTC TGCCCTTTCA ACAGCTACAG GTGACTGATT      720
AAAATTTTAA TTGTGCTTAT TTCAAGCACT TGATTCTGAA AGATGATCAC GATGAGCAGT      780
AAAATCCAGA AGGTAATAAT TTCATACTGT TAATGGATTT TTGGCATCTT GAACATTGCC      840

```

ATAAACCTTT	CAGAATCTGA	GGTAAATCTC	AGATACAGGA	AGTAGCTTGA	AAGAAGACTT	900
ACAGCTGCTG	CTTGGATTTA	GTTACCATAT	GTCTCTATGG	CCACATATTG	TAGCTTTAAT	960
GGATAATATC	GCATTATCCT	GTTGATATTA	TATAAGTATA	TTAGAAGTCA	CAAAGAAAAT	1020
TTCCATAGAA	GGGAATTATG	AAACTTTTTT	TATTTCCAAC	GAGCATACGG	AAGTATGTTT	1080
CATAGCTAAT	TGGATCCCTA	GCCTCAGCAC	AAAAATCTTT	TGTGCCCCGT	GAATACATTT	1140
CTGGAACCTT	GGAGGGCACA	CCCCCATGGT	GGCTGCCCTG	GAGACCTTAG	GTTGGTAATA	1200
TGTAAGGACC	TGAATGTGGA	TGGGCAGAAT	TGGATAAAAG	TCCACGGAAA	AGATGTTACT	1260
CTTGTAATTT	AATAATGTTT	AGCCTGGTGT	CTCTGAAGCC	TATTTCAAAT	AAGCTAGGAG	1320
TTGTGGAGGC	TTTAAGTCCC	ACCAAATAAG	CATAAACATC	CTGATGAAAA	AAGTTTGATG	1380
AATAGTTTGT	TTTTTCTTTT	ATACCAAGCA	TATCTAAAAT	TTTAGAAGAG	TGAAAAGGAA	1440
CCGAGATGGT	GACTGAATCT	TAGGGAAAAA	ATTGTAAATA	GGAAGCCCCT	ATTTGCCTAA	1500
GTATTTTTCT	TGATCCAGTT	AGTATGCTTG	AAATATAACT	TGTCCCAGCA	CCTCATTAAAG	1560
TAGCTTCTTA	GCTGCTCATA	ATTGTTACAG	ATGGAGCATT	CCTAATCCAA	CATCTAAAAT	1620
GCTCCAAAAT	CCAAAACCTT	TTGAGCTTTG	ACATGATGCC	ACAAGTGGAA	AATTCCACAC	1680
CTGACCTCAT	GTGACAGGTC	ACGGTCAGAA	CACAGTCAAA	ATTTTGTTC	ATGCACAAAA	1740
TTACTGAAGA	TATTGTATAA	AATTACTTCA	GGCTATGTGC	ATAAGGTGTA	CAAGAAACAA	1800
ACGAATTTTG	TGTTTAGGCT	TGAGCCTCAT	CCTTAAGATA	CCTCATGTAT	ATGCAAATTT	1860
TCCAAAACCC	AAAAAATTTT	TGAATCTGAA	ATGCTTCTCG	TCCAAATGTT	TCAGGTAAGG	1920
GATATTCAAC	TTGTATTTTT	ATTTTCCTCA	TTCATATACA	GTGTTTTTGA	ATACAGTATT	1980
TTGATCTGCC	TTTAACAAAT	GTTTTCTCAT	TATTTTCAGT	GCCAAAGCTG	TTTGATGGAT	2040
GCTACTTCTA	TTTGTGGGGA	ACCTTCAAAC	ACCATCCAAA	GGACAACCTT	ATTAAGCTCG	2100
TCACTGCAGG	TGGGGGCCAG	ATCCTCAGTA	GAAAGCCCAA	GCCAGACAGT	GACGTGACTC	2160
AGACCATCAA	TACAGTCGCA	TACCATGCGA	GACCCGATTC	TGATCAGCGC	TTCTGCACAC	2220
AGTATATCAT	CTATGAAGAT	TTGTGTAATT	ATCACCAGTA	GAGGGTTCGG	CAGGGCAAAG	2280
TCTGGAAGGC	TCCTTCGAGC	TGGTTTATAG	ACTGTGTGAT	GTCCTTTGAG	TTGCTTCCTC	2340
TTGACAGCTG	AATATTATAC	CAGATGAACA	TTTCAAATTG	AATTTGCACG	GTTTGTGAGA	2400
GCCCAGTCAT	TGTAATGTTT	TTAATGTTCA	CATTTTTTACA	AATAGGTAGA	GTCATTGATA	2460
TTTGTCTTTG	AATCAAAAAA	AAAAAAAAAA	AGTCTAATGC	CAGATTAGGA	ATTCATGTTG	2520
TGTTTACCAT	TTAGAAGCTG	GGATTGCTTT	TAAAGGTTTT	TCTTTTTTAA	ATTGGCATGT	2580
TTTTGATTTA	TCATGTCTTT	CTATTCAGAT	TATTGGGTAT	CAAAGATTAA	TGAGGACACC	2640
AGAATCTTGG	TTAAATAGAC	AAGTGGTATC	ATTACTGTTT	GAGTCTTTTA	ATATTCTCCA	2700
TACCTGCCAC	CAGTGAAAAA	ACTTGCCTTT	TTTTTTTTTT	TTTTTTTAGT	AAACAGAATA	2760
TTATCAAACA	ATTTATTTTG	GCTTTATTGA	AAAAAGAGTA	TTTGGTCTAA	ATGTGCCACC	2820

ATAGGTGTTA AATTCTCCTA TCTGCATTTG TCTTTATCCT ATATTGTGTT CATTTCCTTT	2880
CTTAATAATT TACTTTGTTG TGTGTTTCTA CACTTTCATC CCTGTTTTTT ATCTTGATA	2940
TCATCAGGAA ATTGTGATTT AATCATTAAC ATTGGTTTTT TTGTGTGTGT GGTAAAAATC	3000
AACACTAGGC TCATGGTACA TATTTTTATT CTGTACATTT GCTTGTAAC ATCAATTTGT	3060
AACTCTGTTT ATCTACTACA TGTGTATATA TACTTAGAGC ATTTTCTCTA ACACATTTTA	3120
ATGTTAGTAT TTTTAAAG GTCTGACCAG TCTAGCAAAT TGTCAGTCCA ACGTCATTAC	3180
TTTAAATTAA GAAGCAGTCT TCTTCTGGTA AACCTTGTTG GTATTTGTAA AATAATTTTG	3240
AAGGTCTTAA TTTCTTCCTT TGTAAGGA AAAGGTTTTT TTTAAAGTTT TTAGGTTGGC	3300
ATGGAGGCAG AAGTTGGTGA TTACTTGATT TACAACAGAT TTTTCCAGA TCATACAAAA	3360
GGCCATACAG TAAGTATAGA AGTAGGTATG GGGAGGGCTT ACTAATATCA AATAGGCAAG	3420
GCCTTAGTGA GTGGGCAGGA TACCACTTGA GAGTGGCCAG ATCTGGGGAG GTTACTCTGC	3480
TCTGGGTGCT CTCATTCATG AATCGACAAG GATACATTAG ATTATTTTGA AACATTTTTT	3540
TAAGAAGCAG AATTCTTTAA TAATTCCTTC CTAGACATTG AATATACTTA TAAAATTAAA	3600
GACTTGGGGA AGGAGACACT GAGAGACTTG CCAGTTTGGT TCCTCATGAA CAAAGAGGA	3660
CAGTTTGATA ACTACCAGAA TAGAATATCC CTAGTTTTAA AATAGTGAGA ATCTCTGAAG	3720
TTCATCAACA TCTTAAGATG CACTTACTTG AAAGTTTGAG ATTCTGTTTA TCATTTGAAA	3780
ACACATTTTG CTTTAATTCT TTCTTTGACA TGTGTTTTT TCATATCAAG AAATATATGA	3840
ACAAAATAAT AACCTTTTGA CCCTGACCTT GCTGGGTGAA TTAGCTCTGA AACACTCTCT	3900
ACAACCAGTA ATGCATTTGT CCCACATTTT ATTCTGATAG AAAATGAACA CCATAGCACC	3960
AAACAAAAAT CCGAGGCGTT AGATAATGTC TGGATTAAAT AATTTAAGAC TCTCTAGGAT	4020
TTTGGTTGTC ATTTTTTATT TATAACAGAC TTTAAGTCAC TTTCTGTTGC CTCATAGGTC	4080
ACATTTTAGA CAGGTTTGTG TCTGTTCCCT GCATCTGAAT TCCTGATTGT AAAGACACCT	4140
ATGAGGTCTC TTAGTTTTTG TCATTCATTT TCTTGGTTTA TCACCCCTCC CTTCTTTTTG	4200
TTGTTTTTCC CTGACTGTTA AGCAGTTTCA TCTTTGCTTT TGTTAAATAT TTGACAGCAG	4260
TTAGTTTGTG TTAAGCTCTT GAAACTTGTG ATTGTACTTT CTGTGTAGAT ATACATGTAA	4320
TTATTTTTTA TTTT	4334

CLAIMS:

1. A nucleic acid segment comprising an isolated DNA sequence that encodes a BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide or peptide.

5

2. The nucleic acid segment of claim 1, comprising an isolated DNA sequence that encodes a BARD1 protein, polypeptide, peptide or mutant thereof.

10 3. The nucleic acid segment of claim 2, comprising an isolated DNA sequence that encodes a BARD1 protein characterized as:

(a) being between about 752 and about 777 amino acids in length;

15 (b) comprising an amino-terminal RING motif or domain that mediates the association of BARD1 with the protein BRCA1;

(c) containing ankyrin repeats that are not required for binding to the protein BRCA1;

20 (d) comprising carboxy-terminal BRCT domains that are homologous to the carboxy-terminal sequences of the protein BRCA1;

(e) being encoded by sequences on human chromosome 2q; and

25

(f) binding to the amino-terminal region of the protein BRCA1.

30 4. The nucleic acid segment of claim 2 or 3, comprising an isolated DNA sequence that encodes an isolated BARD1 domain.



5. The nucleic acid segment of claim 4, comprising an isolated DNA sequence that encodes an isolated BARD1 ankyrin repeat, BARD1 BRCT-like, BARD1 RING motif or BARD1 BRCA1-binding domain.

5

6. The nucleic acid segment of any one of claims 2 to 5, comprising an isolated DNA sequence that encodes a wild type BARD1 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from the sequence of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 or SEQ ID NO:39.

10

7. The nucleic acid segment of any one of claims 2 to 6, comprising an isolated DNA sequence that encodes a BARD1 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:2.

15

8. The nucleic acid segment of any one of claims 2 to 6, comprising an isolated DNA sequence that encodes a BARD1 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from the sequence of SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 or SEQ ID NO:39.

20

9. The nucleic acid segment of any one of claims 2 to 6, comprising an isolated DNA sequence that includes a contiguous nucleic acid sequence from between position 75 and position 2406 of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38, or from between position 75 and position 2385 of SEQ ID NO:26.

25

30

10. The nucleic acid segment of any one of claims 2 to 6, comprising an isolated DNA sequence that encodes a full length wild type BARD1 protein having the contiguous amino acid sequence of SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 or SEQ ID NO:39.

5

11. The nucleic acid segment of claim 10, having the DNA sequence of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38.

10

12. The nucleic acid segment of any one of claims 2 to 5, comprising an isolated DNA sequence that encodes a mutant BARD1 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37.

15

13. The nucleic acid segment of claim 12, comprising an isolated DNA sequence that includes a contiguous nucleic acid sequence from between position 75 and position 2406 of SEQ ID NO:32, SEQ ID NO:34 or SEQ ID NO:36.

20

14. The nucleic acid segment of claim 12, comprising an isolated DNA sequence that encodes a full length mutant BARD1 protein having the contiguous amino acid sequence of SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37.

25

15. The nucleic acid segment of claim 14, having the DNA sequence of SEQ ID NO:32, SEQ ID NO:34 or SEQ ID NO:36.

30

16. The nucleic acid segment of claim 12, comprising an isolated DNA sequence that encodes a mutant BARD1 peptide of from about six to about thirty amino acids in length, the peptide including at least one amino acid that is different to the amino acid in the corresponding position within the wild type BARD1 protein sequence, the difference being a mutation that is indicative of a malignant phenotype.

17. The nucleic acid segment of claim 1, comprising an isolated DNA sequence characterized as:

(a) a B123 DNA sequence encoding a B123 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:19;

(b) a BE2 DNA sequence encoding a BE2 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:41;

(c) a BE14 DNA sequence encoding a BE14 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:43;

(d) a BE31 DNA sequence encoding a BE31 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:45; or

(e) a BE445 DNA sequence encoding a BE445 protein or peptide that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:47.

18. The nucleic acid segment of claim 17, wherein said isolated DNA sequence is characterized as:

- 5 (a) a B123 DNA sequence that includes a contiguous nucleic acid sequence from between position 46 and position 864 of SEQ ID NO:17;
- (b) a BE2 DNA sequence that includes a contiguous nucleic acid sequence from between position 37 and position 819 of SEQ ID NO:40;
- 10 (c) a BE14 DNA sequence that includes a contiguous nucleic acid sequence from between position 1 and position 666 of SEQ ID NO:42;
- (d) a BE31 DNA sequence that includes a contiguous nucleic acid sequence from between position 1 and position 693 of SEQ ID NO:44; or
- 15 (e) a BE445 DNA sequence that includes a contiguous nucleic acid sequence from between position 1 and position 816 of SEQ ID NO:46.

20 19. The nucleic acid segment of claim 18, characterized as:

- (a) a B123 DNA sequence having the contiguous DNA sequence of SEQ ID NO:17;
- (b) a BE2 DNA sequence having the contiguous DNA sequence of SEQ ID NO:40;
- 25 (c) a BE14 DNA sequence having the contiguous DNA sequence of SEQ ID NO:42;
- (d) a BE31 DNA sequence having the contiguous DNA sequence of SEQ ID NO:44; or
- 30 (e) a BE445 DNA sequence having the contiguous DNA sequence of SEQ ID NO:46.

20. The nucleic acid segment of any preceding claim, wherein said nucleic acid segment comprises a first DNA coding region that encodes a first protein or peptide selected from  
5 BARD1, B123, BE2, BE14, BE31 or BE445 and a second DNA coding region that encodes a second, distinct selected protein or peptide.

21. The nucleic acid segment of claim 20, wherein said second DNA coding region encodes  
10 a selected tumor suppressor protein or peptide.

22. The nucleic acid segment of claim 20 or 21, wherein said first DNA coding region is operatively linked in frame to said second DNA coding region, said first and second DNA  
15 coding regions encoding a fusion protein.

23. A nucleic acid segment characterized as:

- 20 (a) a nucleic acid segment comprising a sequence region that consists of at least about 20 contiguous nucleotides that have the same sequence as, or are complementary to, about 20 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID  
25 NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID  
30 NO:129 or SEQ ID NO:130; or

- (b) a nucleic acid segment of from about 20 to about 20,000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130; or the complements thereof, under standard hybridization conditions.

24. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region of at least about 20 contiguous nucleotides from SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130; or the complements thereof.

25. The nucleic acid segment of claim 23, wherein the segment hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, or the complements thereof, under standard hybridization conditions.

26. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region of at least about 20 contiguous nucleotides from SEQ ID NO:1, or the complement thereof; or  
5 wherein the segment hybridizes to the nucleic acid segment of SEQ ID NO:1, or the complement thereof, under standard hybridization conditions.

27. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region  
10 of at least about 20 contiguous nucleotides from SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38, or the complements thereof; or wherein the segment hybridizes to the nucleic acid segment of SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38,  
15 or the complements thereof, under standard hybridization conditions.

28. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region of at least about 20 contiguous nucleotides from SEQ ID NO:32, SEQ ID NO:34 or SEQ ID NO:36, or the complements thereof; or wherein the segment hybridizes to the nucleic acid  
20 segment of SEQ ID NO:32, SEQ ID NO:34 or SEQ ID NO:36, or the complements thereof, under standard hybridization conditions.

29. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region  
25 of at least about 20 contiguous nucleotides from SEQ ID NO:17, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46, or the complements thereof; or wherein the segment hybridizes to the nucleic acid segment of SEQ ID NO:17, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46, or the complements thereof, under standard hybridization  
30 conditions.

30. The nucleic acid segment of claim 23, wherein the segment comprises a sequence region of at least about 20 contiguous nucleotides from SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 or SEQ ID NO:18, or the complements thereof; or wherein the segment hybridizes to the nucleic acid segment of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 or SEQ ID NO:18, or the complements thereof, under standard hybridization conditions.

31. The nucleic acid segment of claim 24, wherein the segment comprises a sequence region of at least about 25, about 30, about 50, about 100 or about 500 contiguous nucleotides from SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 or SEQ ID NO:46; or the complements thereof.

32. The nucleic acid segment of claim 25, wherein the hybridizing segment is about 30, about 50, about 100, about 500, about 1,000, about 3,000, about 5,000, about 10,000 or about 15,000 nucleotides in length.

33. The nucleic acid segment of claim 31, wherein the segment comprises a sequence region that consists of about 2531 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:28, SEQ ID NO:30 or SEQ ID NO:38, or of about 2510 contiguous nucleotides of SEQ ID NO:26, or the complements thereof.

34. The nucleic acid segment of claim 31, wherein the segment comprises a sequence region that consists of about 2531 contiguous nucleotides of SEQ ID NO:32, SEQ ID NO:34 or SEQ ID NO:36, or the complements thereof.



35. The nucleic acid segment of claim 31, wherein the segment comprises a sequence region that consists of about 938 contiguous nucleotides of SEQ ID NO:17, about 1083 contiguous  
5 nucleotides of SEQ ID NO:40, about 1326 contiguous nucleotides of SEQ ID NO:42, about 834 contiguous nucleotides of SEQ ID NO:44 or about 898 contiguous nucleotides of SEQ ID NO:46, or the complements thereof.

10 36. The nucleic acid segment of any one of claims 23 to 35, further defined as a DNA segment.

15 37. The nucleic acid segment of any one of claims 23 to 35, further defined as a RNA segment.

20 38. The nucleic acid segment of any preceding claim, operatively positioned under the control of a promoter.

25 39. The nucleic acid segment of any preceding claim, comprised within a recombinant vector.

30 40. The nucleic acid segment of any one of claims 23-37, further comprising a second sequence region of at least about 20 contiguous nucleotides that have the same sequence as, or are complementary to, SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123,

SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130, said sequence region and said second sequence region from spatially distant regions within SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26; SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34; SEQ ID NO:36; SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129 or SEQ ID NO:130..

41. A nucleic acid segment in accordance with any one of claims 1 to 22 for use in the preparation of a recombinant BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, mutant or fusion protein thereof.

42. Use of a nucleic acid segment in accordance with any one of claims 1 to 22 in the preparation of a recombinant BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, mutant or fusion protein thereof.

43. A composition comprising at least a first nucleic acid segment in accordance with any one of claims 1 to 40 for use in the preparation of a composition for use in detecting a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence.

44. Use of a composition comprising at least a first nucleic acid segment in accordance with any one of claims 1 to 40 in the preparation of a composition for use in detecting a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence.

45. A nucleic acid segment in accordance with any one of claims 2 to 11 for use in the preparation of a wild type BARD1 composition for use in detecting or purifying a BRCA1 protein.

46. Use of a nucleic acid segment in accordance with any one of claims 2 to 11 in the preparation of a wild type BARD1 composition for use in detecting or purifying a BRCA1 protein.

47. A composition comprising at least a first nucleic acid segment in accordance with any one of claims 1 to 40 for use in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer.

48. Use of a composition comprising at least a first nucleic acid segment in accordance with any one of claims 1 to 40 in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer.

49. A method of using a nucleic acid segment that comprises an isolated BARD1, B123, BE2, BE14, BE31 or BE445 DNA sequence, the method comprising expressing said nucleic acid segment in a recombinant host cell to prepare a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide expression product in said cell.

50. A method for detecting BARD1, B123, BE2, BE14, BE31 or BE445 in a sample, comprising contacting sample nucleic acids from a sample suspected of containing BARD1, B123, BE2, BE14, BE31 or BE445 with a nucleic acid segment that encodes a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide, respectively, under conditions effective to allow hybridization of substantially complementary nucleic acids, and detecting the hybridized complementary nucleic acids thus formed.

51. A method of detecting a BRCA1 protein, comprising contacting a sample suspected of containing a BRCA1 protein with a BRCA1-binding protein selected from a BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide or fusion protein, under conditions effective to allow the formation of BRCA1-BRCA1-binding protein complexes, and detecting the BRCA1-BRCA1-binding protein complexes so formed.
52. A method of purifying a BRCA1 protein, comprising contacting a composition comprising a BRCA1 protein with a BRCA1-binding protein selected from a BARD1, B123, BE2, BE14, BE31 or BE445 protein, peptide or fusion protein, under conditions effective to allow the formation of BRCA1-BRCA1-binding protein complexes, and obtaining the BRCA1 protein from said BRCA1-BRCA1-binding protein complexes.
53. A method for identifying a patient having or at risk for developing cancer, comprising determining the type or amount of BARD1, B123, BE2, BE14, BE31 or BE445 present within a biological sample from said patient, wherein the presence of a BARD1, B123, BE2, BE14, BE31 or BE445 mutant or an altered amount of wild type BARD1, B123, BE2, BE14, BE31 or BE445 in comparison to a sample from a normal subject, is indicative of a patient having or at risk for developing cancer.
54. A recombinant host cell comprising a nucleic acid segment in accordance with any one of claims 1 to 40.
55. A composition comprising an isolated BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, domain, mutant or fusion protein thereof.

56. The composition of claim 55, comprising an isolated BARD1 protein, polypeptide, peptide, domain or fusion protein thereof that includes a contiguous amino acid sequence of at least about six amino acids from SEQ ID NO:2, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 or SEQ ID NO:39.

57. A BARD1 protein, polypeptide, peptide, domain, mutant or fusion protein thereof for use in the preparation of an anti-BARD1 antibody.

58. Use of a BARD1 protein, polypeptide, peptide, domain, mutant or fusion protein thereof in the preparation of an anti-BARD1 antibody.

59. A BARD1 protein, polypeptide, peptide, domain or fusion protein thereof for use in the detection or purification of a BRCA1 protein.

60. Use of a BARD1 protein, polypeptide, peptide, domain or fusion protein thereof in the detection or purification of a BRCA1 protein.

61. A BARD1 protein, polypeptide, peptide, domain or fusion protein thereof for use in the identification of a binding protein agonist or antagonist that alters the binding of BARD1 to BRCA1 or that alters a biological activity of a BRCA1-BARD1 complex.

62. Use of a BARD1 protein, polypeptide, peptide, domain or fusion protein thereof in the identification of a binding protein agonist or antagonist that alters the binding of BARD1 to BRCA1 or that alters a biological activity of a BRCA1-BARD1 complex.

63. A method for identifying a binding protein agonist or antagonist, comprising contacting a composition comprising BRCA1 and either BARD1, B123, BE2, BE14, BE31 or BE445, with a candidate substance and identifying a candidate substance that alters the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or that alters a biological activity of a complex comprising BRCA1 and either BARD1, B123, BE2, BE14, BE31 or BE445.

64. An antibody having immunospecificity for a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide.

65. An anti-BARD1 antibody for use in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer.

66. Use of an anti-BARD1 antibody in the preparation of a diagnostic formulation for use in identifying a patient having or at risk for developing cancer.

67. A method for detecting BARD1, B123, BE2, BE14, BE31 or BE445 in a sample, comprising contacting a sample suspected of containing BARD1, B123, BE2, BE14, BE31 or BE445 with a first antibody that binds to a BARD1, B123, BE2, BE14, BE31 or BE445 protein or peptide, respectively, under conditions effective to allow the formation of immune complexes, and detecting the immune complexes thus formed.

68. A method of identifying a candidate tumor suppressor gene or oncogene, comprising the steps of:

- (a) obtaining a first DNA segment comprising a candidate gene; said first DNA segment expressing a first fusion protein comprising a transcriptional

transactivating domain operatively attached to the candidate protein encoded by said candidate gene;

- 5 (b) obtaining a second DNA segment that expresses a second fusion protein comprising a BRCA1 or BARD1 RING domain operatively attached to a DNA binding domain that binds to a defined nucleic acid sequence;
- (c) providing said first and second DNA segments to a eukaryotic host cell that comprises a marker gene operatively positioned downstream of said defined  
10 nucleic acid sequence; and
- (d) identifying a eukaryotic host cell that expresses said marker gene, thereby identifying said candidate gene as a candidate tumor suppressor gene or  
15 oncogene.

69. The method of claim 68, wherein said second DNA segment in step (b) expresses a second fusion protein comprising a BRCA1 RING domain.

20 70. The method of claim 68, wherein said second DNA segment in step (b) expresses a second fusion protein comprising a BARD1 RING domain.

25 71. The method of claim 68, wherein said method further comprises isolating the candidate tumor suppressor gene or oncogene identified in step (d) from said first DNA segment.

30 72. The method of claim 68, wherein said first fusion protein comprises a GAL4 or a VP16 transcriptional transactivating domain.

73. The method of claim 72, wherein said second fusion protein comprises a GAL4 DNA binding domain and wherein said defined nucleic acid sequence comprises a GAL4 binding domain recognition sequence.

5

74. The method of claim 68, wherein said eukaryotic host cell is a yeast host cell.

75. The method of claim 68, wherein said eukaryotic host cell is a mammalian host cell.

10

76. The method of claim 68, wherein said method comprises the steps of:

15

(a) obtaining a plurality of first DNA segments comprising a plurality of candidate tumor suppressor genes or oncogenes;

(b) obtaining multiple copies of said second DNA segment;

20

(c) providing said plurality of first DNA segments and multiple copies of said second DNA segments to a population of said eukaryotic host cells in an amount sufficient to provide about one first DNA segment and at least about one second DNA segment to each host cell in said population;

25

(d) culturing said population of cells under conditions and for a period of time effective to allow marker gene expression; and

(e) detecting a host cell from said population that expresses said marker gene, thereby identifying the presence in said cell of a first DNA segment that comprises a candidate tumor suppressor gene or oncogene.



1/8

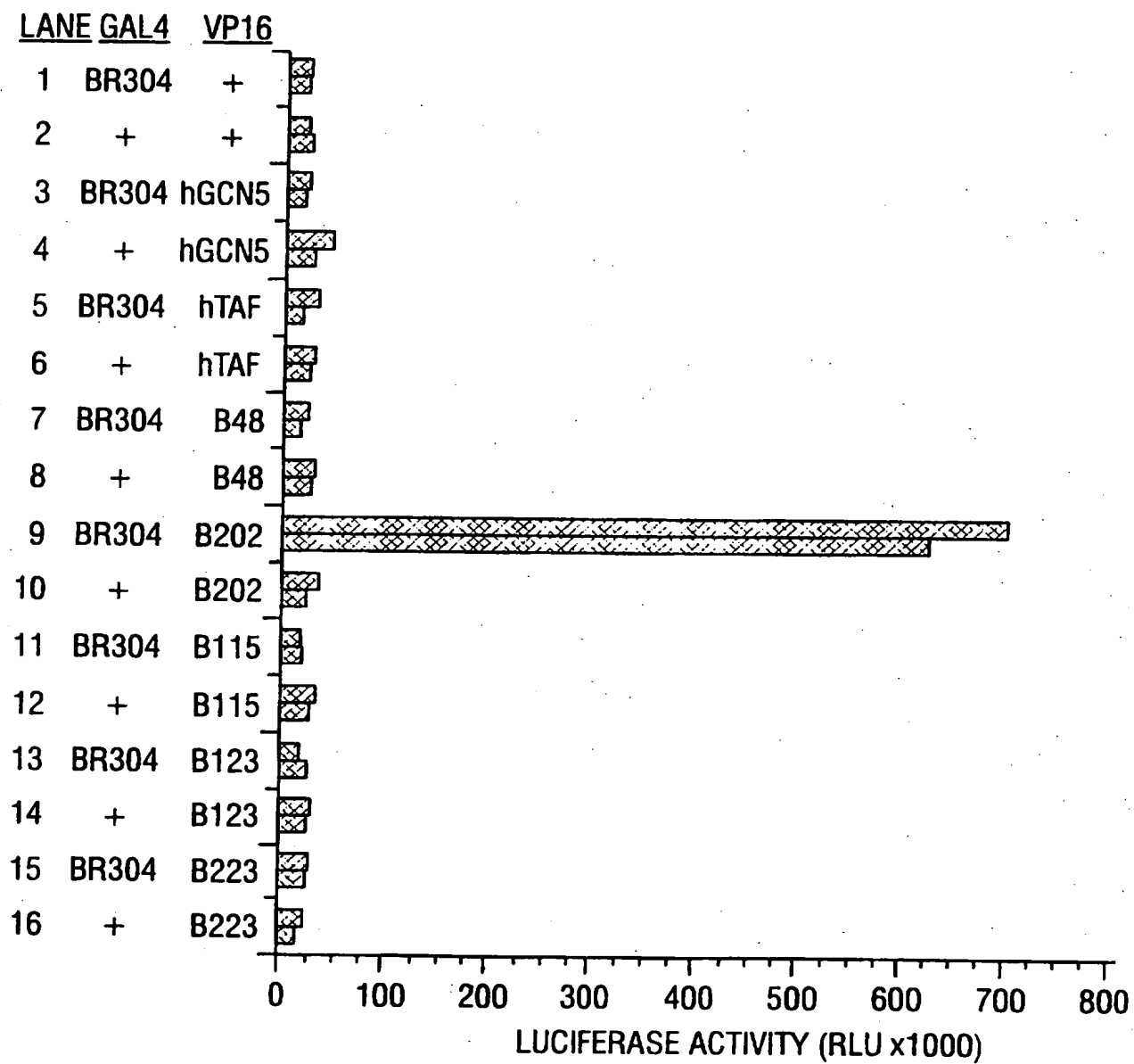


FIG. 1

2/8

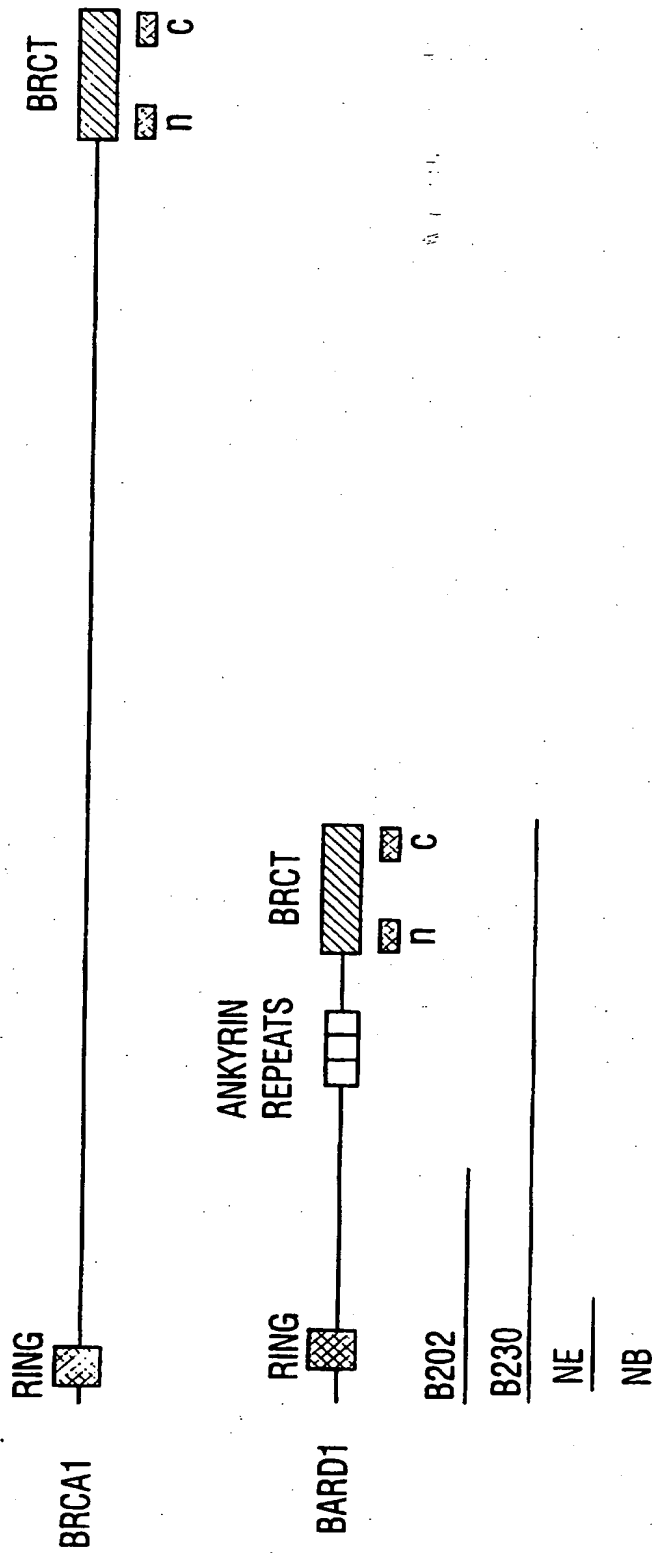


FIG. 2

3/8

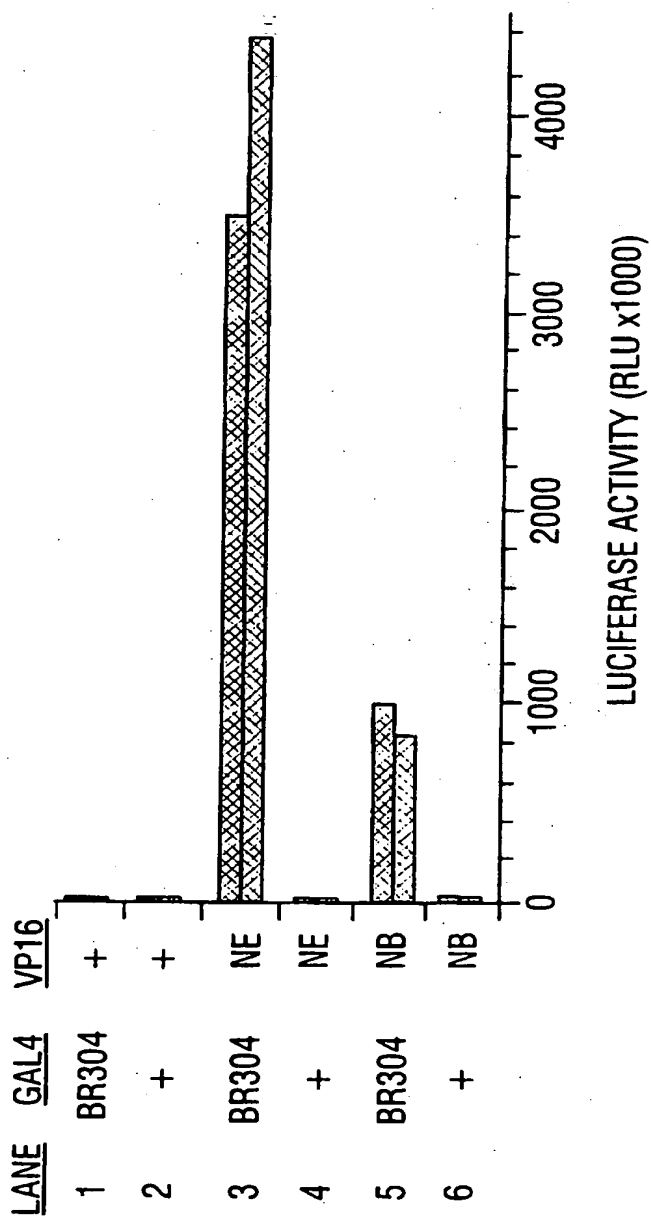


FIG. 3

4/8

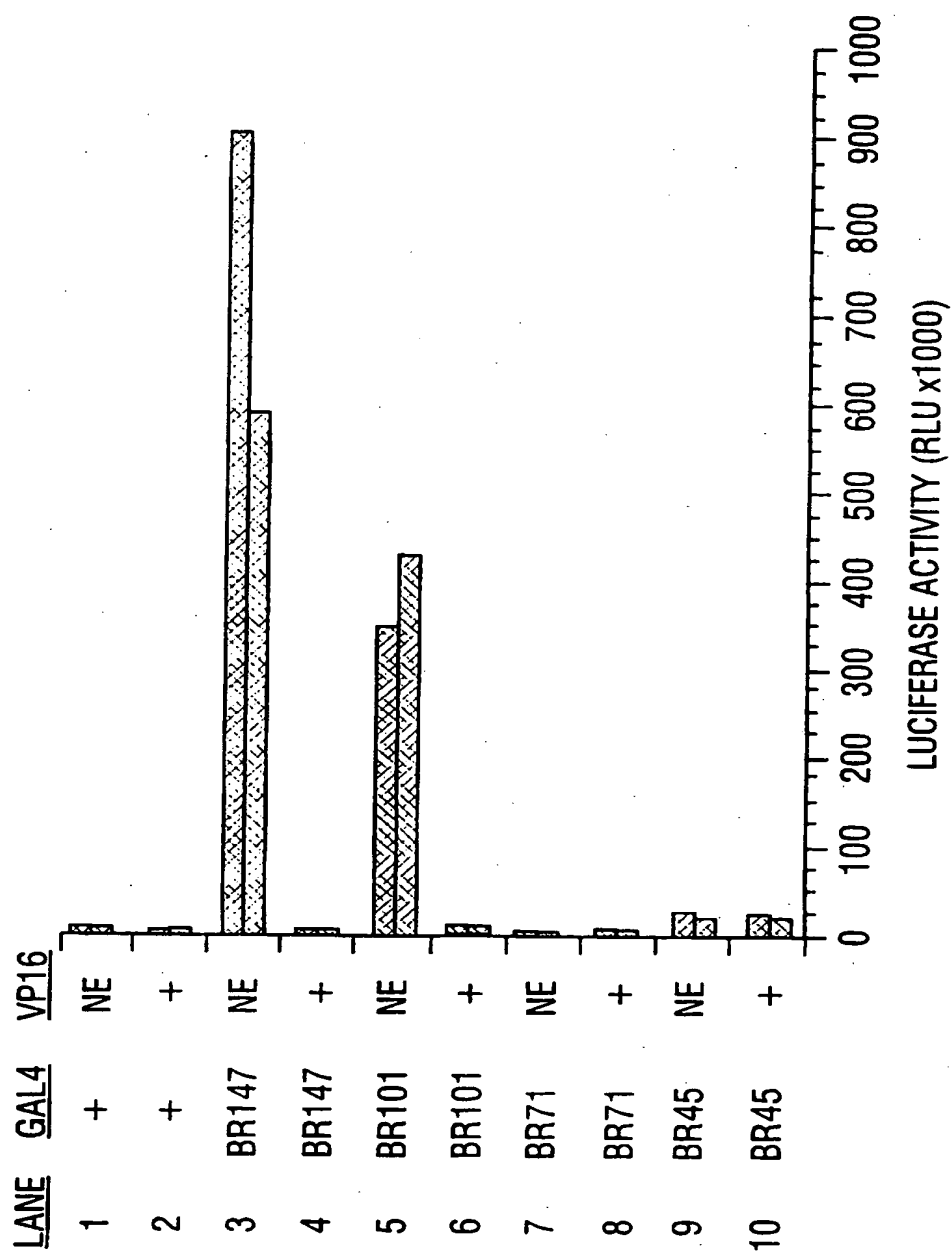


FIG. 4A

5/8

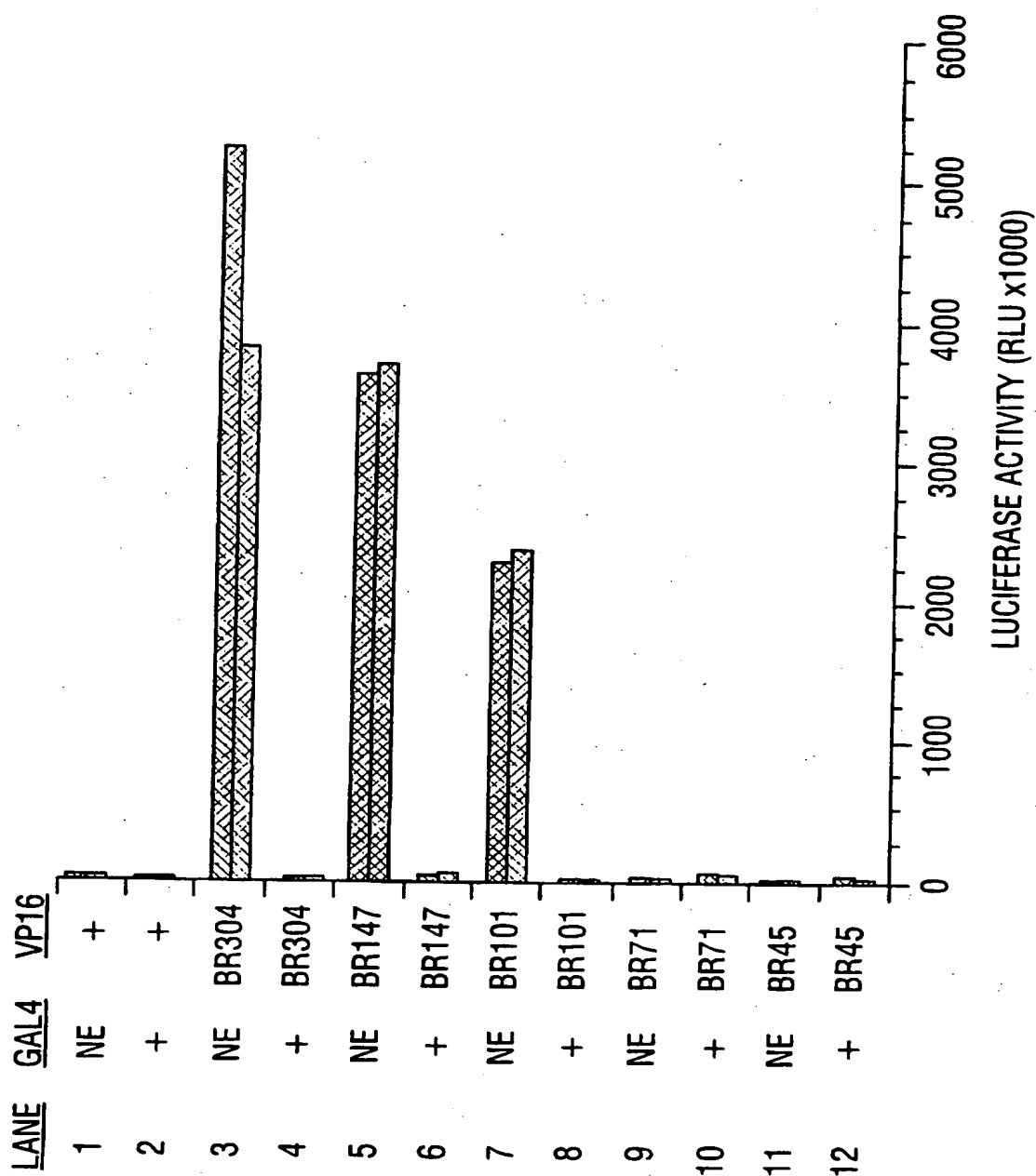


FIG. 4B

6/8

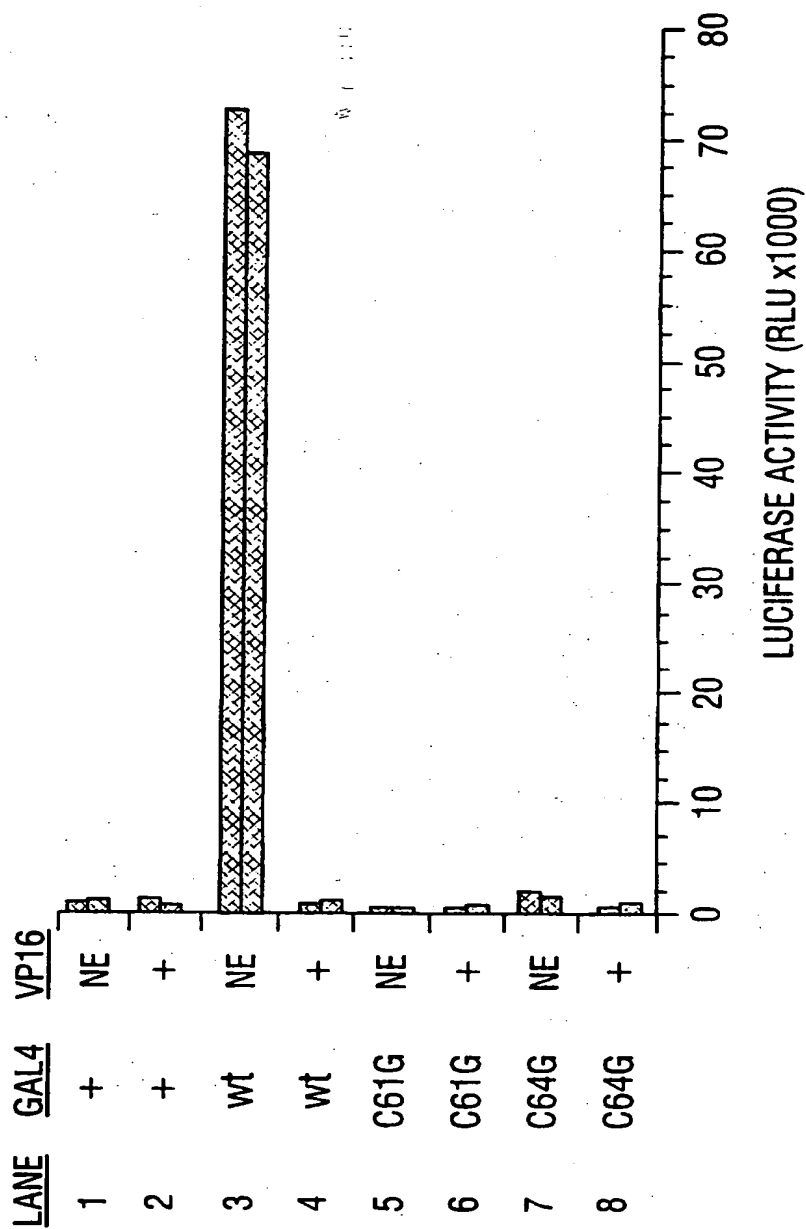


FIG. 5A

7/8

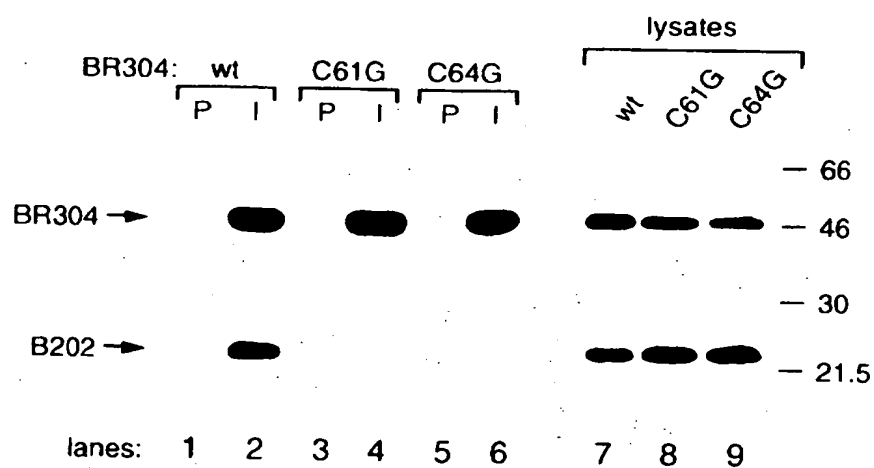


FIG. 5B

8/8

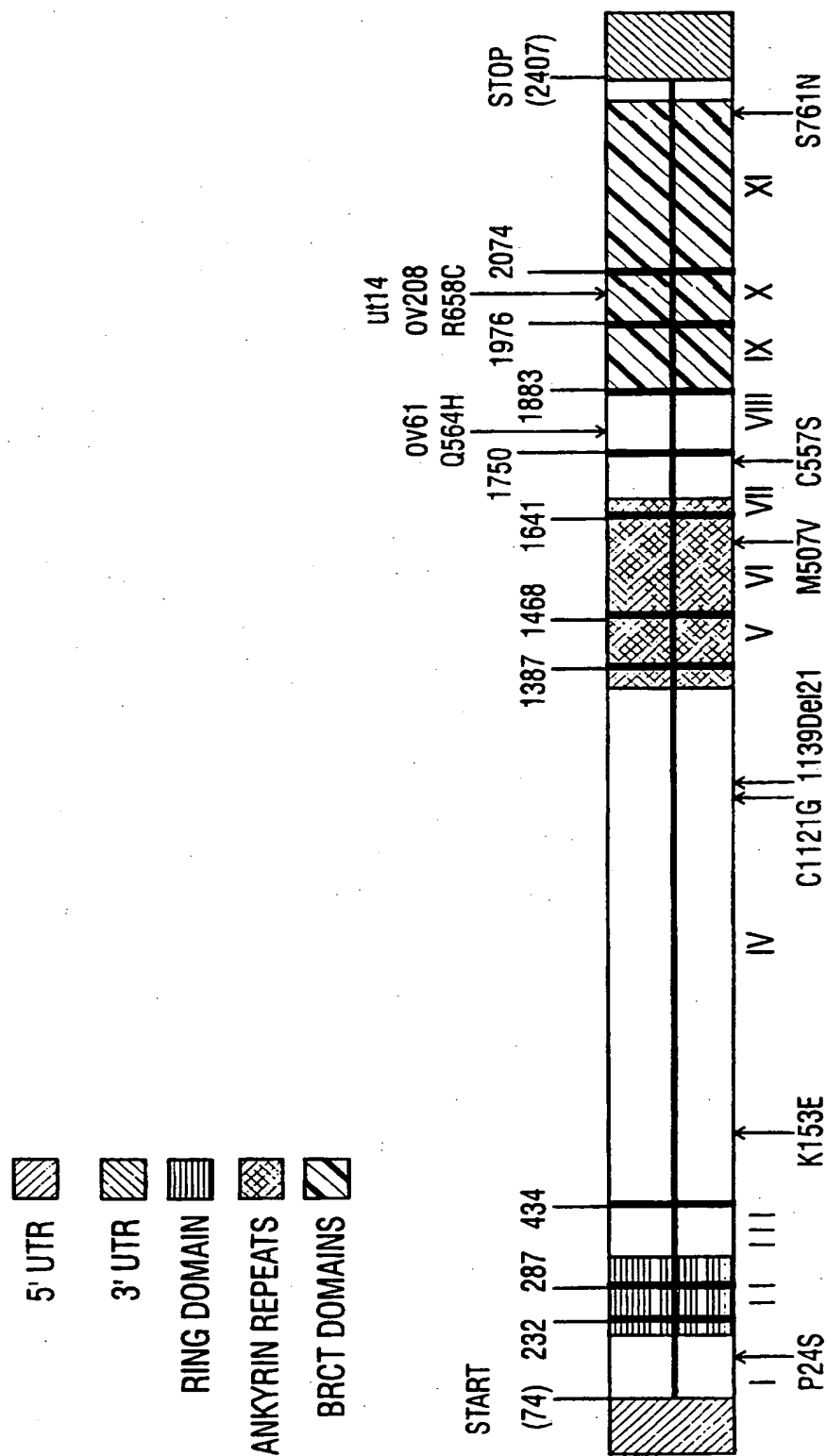


FIG. 6



**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification<sup>6</sup>:</b> <b>C12N 15/12, C07K 14/47, A61K 38/17, G01N 33/68, C12Q 1/68, G01N 33/566, C07K 16/18, C12N 15/62</b>		<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 98/12327</b> <b>(43) International Publication Date:</b> 26 March 1998 (26.03.98)	
<b>(21) International Application Number:</b> PCT/US97/16842 <b>(22) International Filing Date:</b> 19 September 1997 (19.09.97)  <b>(30) Priority Data:</b> 60/025,296 20 September 1996 (20.09.96) US 60/042,611 3 April 1997 (03.04.97) US 60/042,985 4 April 1997 (04.04.97) US  <b>(60) Parent Applications or Grants</b> (63) Related by Continuation US 60/025,296 (CIP) Filed on 20 September 1996 (20.09.96) US 60/042,611 (CIP) Filed on 3 April 1997 (03.04.97) US 60/042,985 (CIP) Filed on 4 April 1997 (04.04.97)  <b>(71) Applicant (for all designated States except US):</b> BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).			<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> BOWCOCK, Anne, M. [GB/US]; 5823 Lupton Drive, Dallas, TX 75225 (US). BAER, Richard [US/US]; 2738 Gold Rush Lane, Carrollton, TX 75007 (US). <b>(74) Agent:</b> HIBLER, David, W.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).  <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>  <b>(88) Date of publication of the international search report:</b> 3 September 1998 (03.09.98)	
<b>(54) Title:</b> COMPOSITIONS AND METHODS COMPRISING BARD1 AND OTHER BRCA1 BINDING PROTEINS				
<b>(57) Abstract</b>  Disclosed are several novel genes, identified in screening assays based upon binding to the breast cancer protein, BRCA1. The currently preferred gene and protein, termed BARD1, is a RING protein that interacts with BRCA1. The genes, proteins and other biological materials provided are envisioned for use in various cancer-related diagnostic and therapeutic methods, particularly those connected with breast, ovarian and uterine cancer.				

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

# INTERNATIONAL SEARCH REPORT

Internal Application No  
PCT/US 98/16842

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 A61K38/17 G01N33/68 C12Q1/68  
G01N33/566 C07K16/18 C12N15/62

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K A61K G01N C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	L.C. WU ET AL.: "Identification of a RING protein that can interact in vivo with the BRCA1 gene product." NATURE GENETICS, vol. 14, December 1996, pages 430-440, XP002050113 see the whole document ---	1-16, 20-28, 31-34, 36-76
P,X	R. SCULLY ET AL.: "Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage" CELL, vol. 90, 8 August 1997, pages 425-435, XP002054690 see the whole document ---	1-16, 20-28, 31-34, 36-68, 70-76
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

7 July 1998.

Date of mailing of the international search report

21.07.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hix, R

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/16842

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,Y	US 5 622 829 A (KING MARY-CLAIRE ET AL) 22 April 1997  see the whole document	1-16, 20-28, 31-34, 36-68, 70-76
P,Y	US 5 654 155 A (MURPHY PATRICIA D ET AL) 5 August 1997  see the whole document	1-16, 20-28, 31-34, 36-68, 70-76
P,Y	WO 97 30108 A (UNIV VANDERBILT ;UNIV WASHINGTON (US)) 21 August 1997  see the whole document	1-16, 20-28, 31-34, 36-68, 70-76
X	E.V. KOONIN ET AL.: "...Functional motifs..." NATURE GENETICS, vol. 13, July 1996, pages 266-267, XP002054691 cited in the application see the whole document	68,71-76
A	S.K. SHARAN ET AL.: "Murine Brcal: sequence and significance for human missense mutations" HUMAN MOLECULAR GENETICS, vol. 4, no. 12, 1995, pages 2275-2278, XP002054692 cited in the application see the whole document	
A	A.J. SAURIN ET AL.: "Does this have a familiar RING?" TRENDS IN BIOLOGICAL SCIENCE, vol. 21, June 1996, pages 208-214, XP004050893 cited in the application see the whole document	
Y	WO 96 05307 A (MYRIAD GENETICS INC ;UNIV UTAH RES FOUND (US); US GOVERNMENT (US)) 22 February 1996 see the whole document	68,71-76
5 Y	WO 96 05308 A (MYRIAD GENETICS INC ;UNIV UTAH RES FOUND (US); US GOVERNMENT (US)) 22 February 1996 see the whole document	68,71-76
3	-/-	

# INTERNATIONAL SEARCH REPORT

Internat. Application No  
PCT/US 96/16842

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 96 05306 A (MYRIAD GENETICS INC ; RECH DU CHUL CENTRE (CA); CANCER INST (JP)) 22 February 1996 see the whole document	68,71-76
P,X	--- J. S. HUMPHREY ET AL.: "Human BRCA1 inhibits growth in yeast: Potential use in diagnostic testing." PROC. NATL. ACAD. SCI. USA, vol. 94, May 1997, pages 5820-5825, XP002067827 see the whole document	68,69, 71-76
P,X	--- WANG H ET AL: "BRCA1 PROTEINS ARE TRANSPORTED TO THE NUCLEUS IN THE ABSENCE OF SERUM AND SPLICE VARIANTS BRCA1A, BRCA1B ARE TYROSINE PHOSPHOPROTEINS THAT ASSOCIATE WITH E2F, CYCLINS AND CYCLIN DEPENDENT KINASES" ONCOGENE, vol. 15, no. 2, 10 July 1997, pages 143-157, XP002050114 see the whole document	68,71-76
P,X	--- A.N.A. MONTEIRO ET AL. : "Evidence for a transcriptional activation function of BRAC1 C-terminal region." PROC. NATL. ACAD. SCI. USA, vol. 93, November 1996, pages 1395-13599, XP002067828 see the whole document	68,71-76
X	--- M.S. CHAPMAN ET AL. : "Transcriptional activation by BRCA1" NATURE, vol. 382, 22 August 1996, pages 678-679, XP002067830 see the whole document	68,71-76
X	--- TAHIRA ET AL.: "Activation of a human c-raf-1 by replacing the N-terminal region with different sequences." NUCLEIC ACID RESEARCH, vol. 15, 1987, pages 4809-4820, XP002070598 see the whole document	23-25, 29,31, 32,36, 37,40
X	--- PRIMIANO ET AL.: "Isolation of cDNAs representing dithiolethione-responsive genes." CARCINOGENESIS, vol. 17, 1996, pages 2297-2303, XP002070599 see the whole document	17
	--- -/--	

# INTERNATIONAL SEARCH REPORT

Term of Application No

PCT/US 97/16842

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>DATABASE EMBL accession number AF007217 14 July 1997 CHANG ET AL.: "A novel thyroid hormone receptor coactivator negatively regulated by the retinoblastoma protein" XP002070604 sequence</p> <p style="text-align: center;">---</p>	<p>17, 23-25, 29,31, 32,36, 37,40</p>
X	<p>SWANSON ET AL.: "A ubiquitin C-terminal hydrolase gene on the proximal short arm of the X chromosome: implications for X-linked retinal disorders." HUMAN MOLECULAR GENETICS, vol. 5, 26 April 1996, pages 533-538, XP002070600 see the whole document</p> <p style="text-align: center;">---</p>	<p>17, 23-25, 29,31, 32,36, 37,40</p>
X	<p>SHANKAR ET AL.: "Macromolecular properties and polymeric structure of canine tracheal mucins." THE BIOCHEMICAL JOURNAL, vol. 276, 1 June 1991, pages 525-532, XP002070601 see the whole document</p> <p style="text-align: center;">---</p>	<p>17</p>
P,X	<p>DATABASE EMBL accession number AF015913 23 August 1997 MARCUS ET AL.: "SKB1Hs, a human homolog of the fission yeast skb1 gene." XP002070605 sequence</p> <p style="text-align: center;">---</p>	<p>23-25, 29,31, 32,36, 37,40</p>
X	<p>CROSS ET AL.: "Purification of CpG islands using a methylated DNA binding column" NAT. GENETICS, vol. 6, 1994, pages 236-244, XP000578157 see the whole document</p> <p style="text-align: center;">---</p>	<p>23-25, 29,31, 32,36, 37,40</p>
X,P	<p>DATABASE EMBL accession number D87462 9 November 1996 NAGASE ET AL.: "Prediction of the coding sequences of unidentified human genes. " XP002070606 sequence</p> <p style="text-align: center;">---</p>	<p>23-25, 29,31, 32,36, 37,40</p>
	-/--	

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US89/16842

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NAGASE ET AL.: "Prediction of the coding sequences of unidentified human genes" DNA RESEARCH, vol. 2, 1995, pages 167-174, XP002037666 see the whole document ---	23-25, 29,31, 32,36, 37,40
X	CLAVERIE ET AL.: "Alu-alert" NATURE, vol. 371, 1994, page 752 XP002070602 see the whole document ---	23-25, 29,31, 32,36, 37,40
X	JURKA ET AL.: "Reconstruction and analysis of human Alu genes" J. MOL. EVOL., vol. 32, 1991, pages 105-121, XP000653048 see the whole document ---	23-25, 29,31, 32,36, 37,40
X	SHORTRIDGE ET AL.: "A human tRNA gene heterocluster encoding threonine, proline and valine tRNAs." GENE, vol. 79, 1989, pages 309-324, XP002070603 see the whole document -----	23-25, 29,31, 32,36, 37,40

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 97/16842

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☒ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.



# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 97/16842

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: Claims: 1 to 67,  
69 and 70 {completely} and Claim 68,  
71-76 {partially},

Nucleic acid segments encoding BARD1, B123, BE2, BE14, BE31 or BE445 protein, polypeptide, peptide, mutant, domain, use of the nucleic acid segments in the preparation of these proteins recombinantly by expressing said nucleic acid in a recombinant host cell, compositions comprising the nucleic acid for the detection or purification of BARD1, B123, BE2, BE14, BE31 or BE445 composition comprising said nucleic acid segment used in the preparation of a diagnostic formulation used in identification of patients at risk of developing cancer, method for detection and purifying BRCA1, B123, BE2, BE14, BE31 or BE445 using a composition comprising BARD1, B123, BE2, BE14, BE31 or BE445, composition comprising isolated BARD1, B123, BE2, BE14, BE31 or BE445, method for identifying a binding protein agonist or antagonist using a composition comprising BRCA1, B123, BE2, BE14, BE31 or BE445 and BARD1, antibodies specific for BARD1, B123, BE2, BE14, BE31 or BE445, anti-BARD1, B123, BE2, BE14, BE31 or BE445 antibodies and a method of identifying a tumour suppressor gene or oncogene using a DNA segment encoding a BARD1 RING domain or BRCA1 RING domain. Nucleic acid segments comprising SEQ ID NOs 9-16 and 18 to TCL52 DNA, TCL163 DNA, B223 DNA, B115 DNA, BAP28 DNA, B48 DNA, B258 DNA, BAP152 DNA, B268 DNA, composition comprising said nucleic acid segment used in the preparation of a diagnostic formulation used in identification of patients at risk of developing cancer, recombinant host cell comprising said nucleic acid segment.

2. Claims: Claims: 68, 71-76 {partially}

Method of identifying a tumour suppressor gene or oncogene using a DNA segment that expresses a fusion protein comprising a BRCA1, as far as not covered by the first invention, ie. as far as the interaction with the BRCA1 gene does not comprise the RING domain at the N-terminal domain.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/16842

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5622829	A	22-04-1997	AU 5566896 A CA 2217668 A EP 0821733 A WO 9633271 A	07-11-1996 24-10-1996 04-02-1998 24-10-1996
US 5654155	A	05-08-1997	AU 1977897 A EP 0820526 A WO 9729213 A US 5750400 A	28-08-1997 28-01-1998 14-08-1997 12-05-1998
WO 9730108	A	21-08-1997	AU 2065397 A	02-09-1997
WO 9605307	A	22-02-1996	US 5747282 A US 5710001 A AU 691958 B AU 3242895 A AU 686004 B AU 3321295 A AU 691331 B AU 3321695 A CA 2196790 A CA 2196795 A CA 2196797 A CN 1159829 A EP 0705902 A EP 0699754 A EP 0705903 A FI 970513 A FI 970514 A FI 970515 A JP 10505742 T NO 970624 A NO 970625 A NO 970626 A WO 9605306 A WO 9605308 A US 5693473 A US 5709999 A US 5753441 A	05-05-1998 20-01-1998 28-05-1998 07-03-1996 29-01-1998 07-03-1996 14-05-1998 07-03-1996 22-02-1996 22-02-1996 22-02-1996 17-09-1997 10-04-1996 06-03-1996 10-04-1996 07-04-1997 07-04-1997 07-04-1997 09-06-1998 14-04-1997 14-04-1997 14-04-1997 22-02-1996 22-02-1996 02-12-1997 20-01-1998 19-05-1998

# INTERNATIONAL SEARCH REPORT

Information on family members

International Application No

PCT/US 16842

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9605308 A	22-02-1996	US 5747282 A	05-05-1998
		US 5710001 A	20-01-1998
		AU 691958 B	28-05-1998
		AU 3242895 A	07-03-1996
		AU 686004 B	29-01-1998
		AU 3321295 A	07-03-1996
		AU 691331 B	14-05-1998
		AU 3321695 A	07-03-1996
		CA 2196790 A	22-02-1996
		CA 2196795 A	22-02-1996
		CA 2196797 A	22-02-1996
		CN 1159829 A	17-09-1997
		EP 0705902 A	10-04-1996
		EP 0699754 A	06-03-1996
		EP 0705903 A	10-04-1996
		FI 970513 A	07-04-1997
		FI 970514 A	07-04-1997
		FI 970515 A	07-04-1997
		JP 10505742 T	09-06-1998
		NO 970624 A	14-04-1997
		NO 970625 A	14-04-1997
		NO 970626 A	14-04-1997
		WO 9605306 A	22-02-1996
		WO 9605307 A	22-02-1996
		US 5693473 A	02-12-1997
		US 5709999 A	20-01-1998
		US 5753441 A	19-05-1998
WO 9605306 A	22-02-1996	US 5693473 A	02-12-1997
		US 5709999 A	20-01-1998
		AU 691958 B	28-05-1998
		AU 3242895 A	07-03-1996
		AU 686004 B	29-01-1998
		AU 3321295 A	07-03-1996
		AU 691331 B	14-05-1998
		AU 3321695 A	07-03-1996
		CA 2196790 A	22-02-1996
		CA 2196795 A	22-02-1996
		CA 2196797 A	22-02-1996
		CN 1159829 A	17-09-1997

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/16842

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9605306 A		EP 0705902 A	10-04-1996
		EP 0699754 A	06-03-1996
		EP 0705903 A	10-04-1996
		FI 970513 A	07-04-1997
		FI 970514 A	07-04-1997
		FI 970515 A	07-04-1997
		JP 10505742 T	09-06-1998
		NO 970624 A	14-04-1997
		NO 970625 A	14-04-1997
		NO 970626 A	14-04-1997
		WO 9605307 A	22-02-1996
		WO 9605308 A	22-02-1996
		US 5747282 A	05-05-1998
		US 5710001 A	20-01-1998
		US 5753441 A	19-05-1998